

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2004, 00:29:03 ; Search time 57 Seconds  
(without alignments)  
2807.166 Million cell updates/sec

Title: US-09-875-519A-22  
Perfect score: 8609  
Sequence: 1 MPTSGPSLLLLLLHPLPA.....KQCQLGAFPTSNVFGCPN 1663

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Watch 100%  
Listing first 45 summaries

Database : PIR 79:\*

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8598	99.9	1663	1 C3HU	complement C3 prec
2	6825	79.3	1663	1 C3RT	complement C3 prec
3	6792.5	78.9	1666	1 C3GP	complement C3 prec
4	6749	78.4	1663	1 C3MS	complement C3 prec
5	4518.5	52.5	1852	2 I50711	complement C3 prec
6	4389.5	51.0	1851	1 C3NU	complement C3 prec
7	4224	49.1	1642	2 I51018	cobra venom factor
8	3452.5	40.1	1620	2 I51339	complement compone
9	3082	35.8	726	2 A27602	complement C3 - ra
10	2490.5	28.9	1673	2 I50806	complement compone
11	2438.5	28.3	1230	2 T30517	complement C3 - Q2 -
12	2148	25.0	1820	2 S21045	complement compone
13	2122	24.6	1683	2 T30885	complement compone
14	1840	21.4	1676	1 C5HU	complement C5 prec
15	1817	21.1	1680	1 C5MS	complement C4 prec
16	1815	21.1	1738	1 A24558	complement C4A pre
17	1803	20.9	1744	1 C4HU	sex-limited protei
18	1733.5	20.1	1735	2 S54784	sex-limited protei
19	1732.5	20.1	1736	2 A21876	sex-limited protei
20	1515.5	17.6	1699	2 T28153	complement compone
21	1465	17.0	1617	2 T28153	complement C4 - ch
22	1017	11.8	1474	1 MAHU	alpha-2-mactroglobu
23	1003	11.7	1472	2 A26122	alpha-2-mactroglobu
24	979	11.4	1507	2 T18544	alpha-2-mactroglobu
25	967	11.2	1476	2 JC5143	alpha-1-mactroglobu
26	944.5	11.0	1500	2 A42210	alpha-1-mactroglobu
27	939	10.9	1477	2 A29952	alpha-1 proteinase
28	936	10.9	1487	2 S15904	alpha-1 proteinase
29	935.5	10.9	1464	2 JC5144	murinoglobulin pre

#### RESULT 1

C3HU

complement C3 precursor [validated] - human

N:Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subunit, C:Species: Homo sapiens (man)

C:Date: 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 09-Jul-2004

C:Accession: A94065; A37999; A92187; A27603; A23435; A45830; B45830; A01257; A01258

R:de Bruijn, M.H.L.; Fey, G.H.

Proc. Natl. Acad. Sci. U.S.A. 82, 708-712, 1985

A:Title: Human complement component C3: cDNA coding sequence and derived primary structure

A:Reference number: A94065; MUID:85140166; PMID:2579379

A:Accession: A94065

A:Molecule type: mRNA

A:Residues: 1-1663 <DEB>

A:Cross-references: UNIPROT:P01024; GB:K02765; NID:G179664; PIDN:AAA85332.1; PID:G17966

R:Vik, D.P.; Amiguet, P.; Moffat, G.J.; Fey, M.; Amiguet-Barras, F.; Wetsel, R.A.; Tack

Biochemistry 30, 1080-1085, 1991

A:Title: Structural features of the human C3 gene: intron/exon organization, transcript

A:Reference number: A37999; MUID:9113687; PMID:1703437

A:Contents: intron/exon structure of gene

A:Accession: A37999

A:Molecule type: DNA

A:Residues: 1-25 <VIK>

A:Cross-references: GB:M63423

A:Note: the authors translated the codon GGT for residue 6 as Leu, CCC for residue 7 as

R:Hugli, T.E.

J. Biol. Chem. 250, 8293-8301, 1975

A:Title: Human anaphylatoxin (C3a) from the third component of complement.

A:Reference number: A92187; MUID:76069169; PMID:1238393

A:Accession: A92187

A:Molecule type: protein

A:Residues: 672-680, 'N', 682-699, 'Q', 701-748 <HUG>

R:Doudaki, M.E.; Becherer, J.D.; Lambiris, J.D.

J. Immunol. 140, 1577-1580, 1988

A:Title: A 34-amino acid peptide of the third component of complement mediates properdin.

A:Reference number: A27603; MUID:88154452; PMID:3279119

A:Accession: A27603

A:Molecule type: protein

A:Residues: 1409-1563 <DAO>

R:Hellman, U.; Eggertsen, G.; Engstrom, A.; Sjoquist, J.

Biochem. J. 230, 353-361, 1985

A:Title: Amino acid sequence of the trypsin-generated C3d fragment from human complement

A:Reference number: A23435; MUID:86025442; PMID:3876831

A:Accession: A23435

A:Molecule type: protein

A:Residues: 1002-1022, 'E', 1014-1303 <HEL>

A:Note: sequence corresponding to residues 1072-1100 was not determined but was taken f

R:Poznansky, M.C.; Clissold, P.M.; Lachmann, P.J.

J. Immunol. 143, 1254-1258, 1989

A:Title: The difference between human C3F and C3S results from a single amino acid chan

3.

A:Reference number: A45830; MUID:89309808; PMID:2473125

alpha-2-mactroglobu  
alpha-2-mactroglobu  
ovostatin precursor  
alpha-2 mactroglobu  
pregnancy zone pro  
alpha-2 mactroglobu  
complement C3 - Af  
hypothetical prote  
hypothetical prote  
complement compone  
alpha-1 proteinase  
alpha-1-inhibitor  
Sip(w7) - mouse (f  
hypothetical prote  
complement C3 - pi  
complement C3, alt

#### ALIGNMENTS

30 934.5 10.9 1495 2 S27001  
31 932 10.8 1503 2 T43166  
32 929.5 10.8 1473 2 A20872  
33 924.5 10.7 1476 2 A41185  
34 912 10.6 1482 2 S13495  
35 891.5 10.4 1451 2 B41185  
36 849 9.9 322 2 A32329  
37 814.5 9.5 1519 2 T27829  
38 805 9.4 1508 2 T27828  
39 621 7.2 231 2 S27041  
40 595 6.9 785 2 A29953  
41 463 5.4 609 2 A41081  
42 457 5.3 594 2 I48771  
43 302.5 3.5 1536 2 E72310  
44 288 3.3 77 2 A01259  
45 249 2.9 73 2 I48284



QY 1381 KNTMILEICTRYGDDATWSILDISNMTGTFAPDITDQLKQANGVDVYISKYELDKAFSD 1440  
Db 1381 KNTMILEICTRYGDDATWSILDISNMTGTFAPDITDQLKQANGVDVYISKYELDKAFSD 1440  
QY 1441 RNTLIYLDKVSHEDDCLAFKQVHQYFNVELIQPGAVKYVAYYNLEESCTRFYHPEKEDG 1500  
Db 1441 RNTLIYLDKVSHEDDCLAFKQVHQYFNVELIQPGAVKYVAYYNLEESCTRFYHPEKEDG 1500  
QY 1501 KLNKLGRDELCAEACNCFIQKSDDKVTLEERLDKACPGVDYVYKTRLVKQLNSDFDE 1560  
Db 1501 KLNKLGRDELCAEACNCFIQKSDDKVTLEERLDKACPGVDYVYKTRLVKQLNSDFDE 1560  
QY 1561 YIMAIQTIKSGSDEVQVQOQRTFISPIKREALKLEEKHYLMWGLSDFWGEKPNLSY 1620  
Db 1561 YIMAIQTIKSGSDEVQVQOQRTFISPIKREALKLEEKHYLMWGLSDFWGEKPNLSY 1620  
QY 1621 IIGKDTWVHWPEDSCQDEENKQOCQDYGATESVWFGCPN 1663  
Db 1621 IIGKDTWVHWPEDSCQDEENKQOCQDYGATESVWFGCPN 1663  
RESULT 2  
C3RT  
Complement C3 precursor - rat  
N:Alternate names: 37K phospholipase A2 inhibitory protein  
N:Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subunit;  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 04-Dec-1992 #sequence revision 07-Oct-1994 #text change 09-Jul-2004  
C:Accession: S15764; A54562; A01260; B35979; A35979; PNO566; A32281; S08692  
R:Misumi, Y.; Sohma, M.; Ikehara, Y.  
Nucleic Acids Res. 18, 2178, 1990  
A:Title: Nucleotide and deduced amino acid sequence of rat complement C3.  
A:Reference number: S15764; MUID:190245672; PMID:2336397  
A:Accession: S15764  
A:Molecule type: mRNA  
A:Residues: 1-1663 <MIS>  
A:CROSS-references: UNIPROT:P01026; EMBL:X52477; NID:G56953; PIDN:CRA36716.1; PID:G56954  
R:Stundstrom, S.A.; Komm, B.S.; Ponce-de-Leon, H.; Yi, Z.; Teuscher, C.; Lyttle, C.R.  
J. Biol. Chem. 264, 16941-16947, 1989  
A:Title: Estrogen regulation of tissue-specific expression of complement C3.  
A:Reference number: A54562; MUID:89380332; PMID:2674144  
A:Accession: A54562  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 'P', 1316-1595 <SUN>  
A:CROSS-references: GB:M29866; NID:9203200; PIDN:AAA40837.1; PID:G554423  
R:Jacobs, J.W.; Rubin, J.S.; Hugli, T.E.; Bogardt, R.A.; Mariz, I.K.; Daniels, J.S.; Dau  
Biochemistry 17, 5031-5038, 1978  
A:Title: Purification, characterization, and amino acid sequence of rat anaphylatoxin (C  
A:Reference number: A01260; MUID:79062262; PMID:309768  
A:Accession: A01260  
A:Molecule type: protein  
A:Residues: 671-703, 'K', 705-720, 'KL', 723-748 <JAC>  
A:Note: three disulfide bonds are present  
R:Suwa, Y.; Kudo, I.; Imaizumi, A.; Okada, M.; Kamimura, T.; Suzuki, Y.; Chang, H.W.; Ha  
Proc. Natl. Acad. Sci. U.S.A. 87, 2395-2399, 1990  
A:Title: Proteinaceous inhibitors of phospholipase A-2 purified from inflammatory sites  
A:Reference number: A35979; MUID:90207203; PMID:2320562  
A:Accession: B35979  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 'X', 998-1005 <SUN>  
A:Accession: A35979  
A:Molecule type: protein  
A:Residues: 'X', 961-962, 'P', 964-969 <SU2>  
R:Nakagawa, H.; Komorita, N.  
Biochem. Biophys. Res. Commun. 194, 1181-1187, 1993  
A:Title: Complement component C3-derived neutrophil chemotactic factors purified from ex  
A:Reference number: PNO566; MUID:93356786; PMID:8352775  
A:Accession: PNO567  
A:Molecule type: protein  
A:Residues: 568-592 <NAK>  
A:Note: amino end of a C3-derived peptide designated exudate neutrophil chemotactic fact

A:Accession: PNO566  
A:Molecule type: protein  
A:Residues: 671-687 <NA2>  
A:Note: amino end of peptide designated neutrophil chemotactic factor 1 and probably id  
R:Kuivaneen, P.C.; Capulong, R.B.; Harkins, R.N.; Desombre, E.R.  
Biochem. Biophys. Res. Commun. 158, 898-905, 1989  
A:Title: The estrogen-responsive 110K and 74K rat uterine secretory proteins are structu  
A:Reference number: A32281; MUID:89149812; PMID:2645873  
A:Accession: A32281  
A:Molecule type: protein  
A:Residues: 25-41 <KUI>  
A:Experimental source: 17beta-estradiol-stimulated uterus of immature rat  
A:Note: the authors treat this 74K uterine secretory protein, identical as far as sequen  
ent  
C:Comment: Complement C3 contains two chains, formed by removal of four residues and 11  
alternative-complement-pathway C3/C5 convertase.  
C:Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.  
C:Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign p  
e classical-complement-pathway C3/C5 convertase. The activity of C3b is regulated by pr  
C:Superfamily: alpha-2-macroglobulin  
C:Keywords: acute phase; chemotaxis; complement alternate pathway; complement pathway;  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-666/Product: complement C3 and C3b beta chain #status predicted <C3BB>  
F:25-666, 671-1663/Product: complement C3 #status predicted <CC3>  
F:25-666, 749-1663/Product: complement C3b #status predicted <C3b>  
F:671-1663/Product: complement C3 alpha chain #status predicted <CC3A>  
F:671-748/Product: C3a anaphylatoxin #status experimental <C3T>  
F:749-1663/Product: complement C3b alpha' chain #status predicted <C3BA>  
F:946-1303/Product: C3dk fragment #status predicted <CDK>  
F:1002-1303/Product: C3d fragment #status predicted <C3D>  
F:1424-1457/Region: properdin binding  
F:558-816, 626-661, 693-720, 694-727, 707-728, 873-1513, 1101-1158, 1358-1489, 1389-1458, 1506-1  
F:748-749/Cleavage site: Arg-Ser (C3 convertase) #status predicted  
F:939, 1617/Binding site: Carbohydrate (Asn) (covalent) #status predicted  
F:1010-1013/Cross-link: thiolester (Cys-Gln) #status predicted  
F:1303-1304/Cleavage site: Arg-Ser (complement factor I) #status predicted  
F:1320-1321/Cleavage site: Arg-Ser (complement factor I) #status predicted  
Query Match 79.3%; Score 6825; DB 1; Length 1663;  
Best Local Similarity 78.1%; Pred. NO. 0;  
Matches 1301; Conservative 171; Mismatches 188; Indels 6; Gaps 5;  
QY 1 MGPTSGPSL--LILLTLPLALGSPMYSIIIPNIRLESEETMVLAEHDAQDVPVTV 58  
Db 1 MGPTSGQLLVLLILLASSILLALGSPMYSIIIPNVLRLESEETFILEHDAQDVPVTV 60  
QY 59 VHDPPGKLVLSSEKTVLTPAHNMGNVFTTIPANREFSKGRNKFVTVQATFGQVVE 118  
Db 61 VQDFL-KKQVLTSEKTVLTGATGHLNRVFIKIPASKEFNADKG-HKYVTVWANFGATVVE 118  
QY 119 KVLVLSQGYLFIQDXTIYTPGSTVLYRIFTVNHKLLPVGRVTVMNIENPEGIPVKOD 178  
Db 119 KAVLVSFQGYLFIQDXTIYTPGSTVLYRIFTVNNLLPVGRVTVMNIENPEGIPVKOD 178  
QY 179 SLSSNQGLVPLSDWIDPELVNMGQKIRAYVENSPPQVFSTEFVKEVYLPSEFVIEP 238  
Db 179 ILSSHNQYGLPLSNWIPELVNMGQKIRAFYHAPKQTFSAEFVKEVYLPSEFVIEP 238  
QY 239 TEKFFYYNKEGLEVTITARELYGKKVGTAVIFGIODGEORISLPESLKRIPEIDGSG 298  
Db 239 TEKFFYIHGPKGLEVSIHARFLYKQNVGTAVIFGVQDQDKKISLAUSLFRVLIEDGSG 298  
QY 299 EYVLGRKVLDCQVNPRAEDLVGKSLYVSATVILHSGSDMWQAERSGIPVTSPIQIHFT 358  
Db 299 EAVLSRKVLMDGVRSPSEALVGKSLYVSATVILHSGSDMWQAERSGIPVTSPIQIHFT 358  
QY 359 KTPKYFKCMPEDLVYFTNPDGSPAYRVFVAVQGEDTVQSLTQDGVAKLSINTHPQSK 418  
Db 359 KTPKFKFPAMPFDLMVFTNPDGSPARVPVVTQGSQ-AQALTQDGVAKLSVNTPNRQ 417  
QY 419 PLISITVTKQBELSEAEQATRMQALPYSTVGNSSNNYLHLSVLRLEPGETLNNVFLLR 478

Db 418 PLUTITVSTKKEGIPDARQATRTWQAPYSTVHNSNNYLHLSVSRVLPKPGDNLNHNFLR 477  
QY 479 MDRAEAKRYTYTILMNGRLLKAGQVRBEGDQVLVPLISITTDIPSPRIVAVYTLI 538  
Db 478 TDAGQBAKRYTYTILMNGRLLKAGQVRBEGDQVLVPLISITTDIPSPRIVAVYTLI 537  
QY 539 GASGQREVVDVSVVVDKDCVSGVSLVYVSGQSDRQVPVGGQMTLKIEGDHARVILVAV 598  
Db 538 GANGQREVVDVSVVVDKDCVSGVSLVYVSGQSDRQVPVGGQMTLKIEGDHARVILVAV 597  
QY 599 DKGVPVFNKKNKLTQSKINDVVEKADIGCTPGSGKDVAGVPSDAGLFTSSGGQOATQRA 658  
Db 598 DKGVPVFNKKNKLTQSKINDVVEKADIGCTPGSGKDVAGVPSDAGLFTSSGGQOATQRA 657  
QY 659 ELQCPQAPARRRSVOLTEKRVMDKVGVK - PKELRKCCEDGMREMPFSCQRTRFISLG 717  
Db 658 DPECAPARRRSVOLTEKRVMDKVGVK - PKELRKCCEDGMREMPFSCQRTRFISLG 717  
QY 718 EACKKVFELDCNVITELRQHARASHGLGARSNLDDEIIAEENIVSRSEPPESLWNVED 777  
Db 718 EACKKVFELDCNVITELRQHARASHGLGARSNLDDEIIAEENIVSRSEPPESLWNVED 777  
QY 778 LKPPKNGISTKLMNIFLKDSTTWELLAVNSDKKGCIVADPEVTVMDQFFDIDRLPY 837  
Db 778 LKPPKNGISTKLMNIFLKDSTTWELLAVNSDKKGCIVADPEVTVMDQFFDIDRLPY 837  
QY 838 SVVRNEQVIRAVLYNRYQNELKVRVLLHNPAPFCSLATTKRHQOITIPPKSSLSVP 897  
Db 838 SVVRNEQVIRAVLYNRYQNELKVRVLLHNPAPFCSLATTKRHQOITIPPKSSLSVP 897  
QY 898 YIVPLKGTQREVVEKAAVYHHFISDGVKRSKVVPEGRMKNVAVRTLDPERLREGV 957  
Db 898 YIVPLKGTQREVVEKAAVYHHFISDGVKRSKVVPEGRMKNVAVRTLDPERLREGV 957  
QY 958 QEDIPADLSQVDPDTESETRIILQGTFFVQMTDEDAVDAERLKHILVTPSGCGEQNMIG 1017  
Db 958 QEDIPADLSQVDPDTESETRIILQGTFFVQMTDEDAVDAERLKHILVTPSGCGEQNMIG 1017  
QY 1018 MPTVTIANVHYLDETQWEKFGLEKQGALELIKGYTQQLAFRQPSAPFAVFRAPSTW 1077  
Db 1018 MPTVTIANVHYLDETQWEKFGLEKQGALELIKGYTQQLAFRQPSAPFAVFRAPSTW 1077  
QY 1078 LTAIVVVKVFLANLIIADSVLCGAVKWLILKQKPDGVQEDAPVTHOEMIGGLRNN 1137  
Db 1078 LTAIVVVKVFLANLIIADSVLCGAVKWLILKQKPDGVQEDAPVTHOEMIGGLRNN 1137  
QY 1138 EKDMALTAVFLISLOEAKDICEQVNSLPGSITKAGDFLEANNVNLQSRSTYTHIHWSAS 1197  
Db 1138 EADVSLTAVFLIALQEARDCBQVNSLPGSITKAGDFLEANNVNLQSRSTYTHIHWSAS 1197  
QY 1198 QMRLKGLPLNKFLTTAKDKRWEDPGKQVNSLPGSITKAGDFLEANNVNLQSRSTYTHIHWSAS 1257  
Db 1198 QMRLKGLPLNKFLTTAKDKRWEDPGKQVNSLPGSITKAGDFLEANNVNLQSRSTYTHIHWSAS 1257  
QY 1258 BORYGGGSGTQATFVFWQALQYQKADPHQELNLDVSLQPSRSKXTHRIHWSAS 1317  
Db 1258 BORYGGGSGTQATFVFWQALQYQKADPHQELNLDVSLQPSRSKXTHRIHWSAS 1317  
QY 1318 LURSETEKNEGTFTVTEGKGOTLSVVTWYHAKADOLTCNKFDLKVTIKPAPETKRP 1377  
Db 1318 LURSETEKNEGTFTVTEGKGOTLSVVTWYHAKADOLTCNKFDLKVTIKPAPETKRP 1377  
QY 1378 QAKVNTMILEITRVRGODATMSILDSMTGTFAPDQVNSLPGSITKAGDFLEANNVNLQSRSTYTHIHWSAS 1437  
Db 1378 QAKVNTMILEITRVRGODATMSILDSMTGTFAPDQVNSLPGSITKAGDFLEANNVNLQSRSTYTHIHWSAS 1437  
QY 1438 FSDRNLTIIYLDKSHSEDDCLAFKVOYFNVELIQGAVKVAAYVNLSESCRFVHPEK 1497  
Db 1438 FSDRNLTIIYLDKSHSEDDCLAFKVOYFNVELIQGAVKVAAYVNLSESCRFVHPEK 1497  
QY 1498 EUGKLNKLRDELCAENBNCIFQSDDKVLEBLRDLKACBPGVDYVYKTLRVLKVLQNSD 1557  
Db 1498 EUGKLNKLRDELCAENBNCIFQSDDKVLEBLRDLKACBPGVDYVYKTLRVLKVLQNSD 1557

## RESULT 3

## C3GP

Complement C3 precursor - guinea pig  
Nomenclature: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subunit;  
C-species: *Canis familiaris* (guinea pig)  
C-Date: 07-Feb-1992 #sequence, revision 07-Oct-1994 #text change 09-Jul-2004  
C-Accession: A37156; S03375; A20342; D20342; C20342; A31222  
J. Auerbach, H.S.; Burger, R.; Dodds, A.; Colten, H.R.  
J. Clin. Invest. 96, 96-106, 1990  
A-Title: Molecular basis of complement C3 deficiency in guinea pigs.  
A-Reference number: A37156; MUID: 90307998; PMID: 1973176  
A-Accession: A37156  
A-Molecule type: mRNA  
A-Residues: 1-1666 <AUE>  
A-Cross-references: UNIPROT: P12387; GB: M34054; NID: G191262; PIDN: AAA37038.1; PID: G305335  
R. Gerard, N.P.; Lively, M.O.; Gerard, C.  
Protein Seq. Data Anal. 1, 473-478, 1988  
A-Title: Amino acid sequence of guinea pig C3a anaphylatoxin.  
A-Reference number: S03375; MUID: 89113342; PMID: 3064079  
A-Accession: S03375  
A-Molecule type: Protein  
A-Residues: 676-730, 'N', 732-752 <GER>  
A-Experimental source: Complement-activated guinea pig serum  
R. Thomas, M.L.; Tack, B.F.  
Biochemistry 22, 942-947, 1983  
A-Title: Identification and alignment of a thiol ester site in the third component of guinea pig C3.  
A-Reference number: A90479; MUID: 83178889; PMID: 6838833  
A-Accession: A20342  
A-Molecule type: Protein  
A-Residues: 676-687 <TH1>  
A-Accession: D20342  
A-Molecule type: Protein  
A-Residues: 993-1012, 1014-1017, 'E', 1019-1030, 'Y' <TH2>  
J. Goldberger, G.; Thomas, M.L.; Tack, B.F.; Williams, J.; Colten, H.R.; Abraham, G.N.  
J. Biol. Chem. 256, 12617-12619, 1981  
A-Title: NH2-terminal structure and cleavage of guinea pig pro-C3, the precursor of the C3a anaphylatoxin.  
A-Reference number: A20342; MUID: 82075767; PMID: 6458605  
A-Accession: C20342  
A-Molecule type: Protein  
A-Residues: 23-38 <GOL>  
C-Comment: Complement C3 contains two chains, formed by removal of four residues and lin-  
nate alternative complement pathways, releases the C3a anaphylatoxin from the amino end of the  
native-complement-pathway C3/C5 convertase.  
C-Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.  
C-Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign pa-  
thogens and is regulated by C3/C5 convertase. The activity of C3b is regulated by pro-  
C3. The major site of synthesis of this plasma protein is the liver.  
C-Superfamily: alpha-2-macroglobulin  
C-Keywords: acute phase; complement alternate pathway; complement pathway; glycoprotein;  
F.1-22/Domain: signal sequence #status predicted <SIG>  
F.23-671/Product: complement C3 and C3b beta chain #status predicted <C3BB>  
F.23-671/Product: complement C3 #status predicted <C3>  
F.23-671, 754-1666/Product: complement C3b #status predicted <C3B>  
F.23-671, 754-1666/Product: complement C3 alpha chain #status predicted <CC3>  
F.676-753/Product: C3a anaphylatoxin #status predicted <C3a>  
F.754-1666/Product: complement C3b alpha' chain #status predicted <C3BA>  
F.951-1308/Product: C3dk fragment #status predicted <CDK>  
F.1007-1308/Product: C3dk fragment #status predicted <CDK>  
F.1429-1461/Region: properdin binding  
F.553-821, 630-666, 698-732, 712-733, 878-1517, 1106-1163, 1363-1493, 1394-1462, 1510-1:  
F.753-754/Cleavage site: Arg-Ser (C3 convertase) #status predicted





A;Residues: 671-1663 <WET>  
R;Domdey, H.; Wiebauer, K.; Kazmaier, M.; Muller, V.; Odink, K.; Fey, G.  
Proc. Natl. Acad. Sci. U.S.A. 79, 7619-7623, 1982  
A;Title: Characterization of the mRNA and cloned cDNA specifying the third component of  
A;Reference number: A93938; MUID:83117730; PMID:6961437  
A;Contents: C3a  
A;Accession: A93938  
A;Molecule type: mRNA  
A;Residues: 671-748 <DOM>  
R;Sottrup-Jensen, L.; Stepanik, T.M.; Kristensen, T.; Lonblad, P.B.; Jones, C.M.; Wierzb  
Proc. Natl. Acad. Sci. U.S.A. 82, 9-13, 1985  
A;Title: Common evolutionary origin of alpha2-macroglobulin and complement components C3  
A;Reference number: A21898; MUID:85113177; PMID:2578664  
A;Accession: A21898  
A;Molecule type: mRNA  
A;Residues: 25-1663 <SOT>  
R;Hamada, J.; Cavanaugh, P.G.; Mikil, K.; Nicolson, G.L.  
Cancer Res. 53, 4418-4423, 1993  
A;Title: A paracrine migration-stimulating factor for metastatic tumor cells secreted by  
A;Reference number: A54561; MUID:93373334; PMID:8364938  
A;Accession: A54561  
A;Molecule type: protein  
A;Residues: 25-41749-760 <HAM>  
R;Experimental source: migration-stimulating factor purified from medium conditioned by  
R;Sato, T.; Hong, M.H.; Jin, C.H.; Ishimi, Y.; Udagawa, N.; Shinki, T.; Abe, E.; Suda, T  
FEBS Lett. 285, 21-24, 1991  
A;Title: The specific production of the third component of complement by osteoblastic ce  
A;Reference number: S16189; MUID:91293304; PMID:2065778  
A;Accession: S16189  
A;Molecule type: protein  
A;Residues: 25-31 <SAT>  
A;Accession: S16189  
A;Status: Preliminary  
A;Molecule type: protein  
A;Residues: 671-677, X', 679-680 <SA2>  
R;Fey, G.; Domdey, H.; Wiebauer, K.; Whitehead, A.S.; Odink, K.  
Springer Semin. Immunopathol. 6, 119-147, 1993  
A;Title: Structure and expression of the C3 gene.  
A;Reference number: I49563; MUID:84045280; PMID:6356427  
A;Accession: I49563  
A;Status: Preliminary  
A;Molecule type: mRNA  
A;Residues: 25-136, Y', 138-240 <FEY>  
A;Cross-references: GB:M35659; NID:G192280; PIDN:AAA37339.1; PID:G192281  
R;Fey, G.H.; Wiebauer, K.; Domdey, H.  
Ann. N. Y. Acad. Sci. 421, 307-312, 1983  
A;Title: Amino acid sequences of mouse complement C3 derived from nucleotide sequences c  
A;Reference number: I49576; MUID:84201365; PMID:6609661  
A;Accession: I49576  
A;Status: Preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 658-761 <RES>  
A;Cross-references: GB:M33032; NID:G192391; PIDN:AAA37378.1; PID:G192392  
A;Comment: Complement C3 contains two chains, formed by removal of four residues and lin  
alternative complement pathways releases the C3a anaphylatoxin from the amino end of b  
native-complement-pathway C3/C5 convertase.  
C;Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.  
C;Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign pa  
e classical-complement-pathway C3/C5 convertase. The activity of C3b is regulated by p  
C;Comment: The major site of synthesis of this plasma protein is the liver.  
A;Genetics:  
A;Introns: 27/2; 90/3  
A;Note: the list of introns may be incomplete  
C;Superfamily: alpha-2-macroglobulin  
C;Keywords: acute phase; complement alternate pathway; complement pathway; glycoprotein;  
F;1-24/Domain: signal sequence #status predicted <SIG>  
F;25-666/Product: complement C3 and C3b beta chain #status predicted <C3BB>  
F;25-666/71-1663/Product: complement C3 #status predicted <CC3>  
F;25-666/749-1663/Product: C3b #status predicted <C3B>  
F;571-1663/Product: complement C3 alpha chain #status predicted <CC3A>  
F;571-748/Product: C3a anaphylatoxin #status predicted <C3T>  
F;749-1663/Product: C3b alpha chain #status predicted <C3BA>  
F;946-1303/Product: C3dk fragment #status predicted <CDK>

F;1002-1303/Product: C3d fragment #status predicted <C3D>

F;1424-1457/Region: properdin binding  
F;559-816,626-661,693-720,694-727,703 conversease) #status predicted  
F;748-749/Cleavage site: Arg-Ser (C3 conversease) #status predicted  
F;939,1617/Binding site: carbohydrate (Aen) #status predicted  
F;1010-1013/Cross-link: thiolester (Cys-Gln) #status predicted  
F;1303-1304/Cleavage site: Arg-Ser (complement factor I) #status predicted  
F;1320-1321/Cleavage site: Arg-Ser (complement factor I) #status predicted

Query Match	78.43;	Score 6749;	DB 1;	Length 1663;
Best Local Similarity	77.13;	Pred. No. 0;		
Matches 1284;	Conservative 179;	Mismatches 197;	Indels 6;	Gaps 5;
QY	1	MGPTSGPSL--LLLLLTHLPLALGSPMYSIITPNIRLSESEETWVLEAHDAQSDVPVTVT	58	
Db	1	MGPASGSQLVLLLLASSPLALGIPMYSIITPNVRLSESEETVLEAHDAQSDIPVTVT	60	
QY	59	VHDPFGKLVLSSEKTVLTTPATNMGNVTTPANREFKSEKGRNKEVTVQATFGTVVE	118	
Db	61	VQDFL-KFQVLTSEKTVLTGASGHLRSVSIKIPASKEFNSDKGKHYTVVAFGETWE	119	
QY	119	KVLVSLQSGYLFQTDKTIYTPGSTVLYXRTFVNHKLLPVGRVTVMVNIENPBGIPVKQD	178	
Db	120	KAMVSPGSGYLFQTDKTIYTPGSTVLYXRTFVNDNLLPVGKTVVILIETPDGIPVKRD	179	
QY	179	SLSSQNLGVLPPLSWDIPELVNMQWKIRAYENSPQOVFTSTFEVKEYVLPSEFVIVEP	238	
Db	180	ILSSNNQHGILPLSWNIPELVNMQWKIRAFYHAPKQIFSAEFVKEYVLPSEFVRVEP	239	
QY	239	TEKFVYIYNEKGLVTTARELYGKVEGTAFVIFGQDGEORISLPESLKRIPIEDGSG	298	
Db	240	TETFYIIDDPNGLEVSIIAKFLYGNVDGTAFAVFGVDGDKKLSLAHSITRVVIEDGVG	299	
QY	299	EVVLSRKVLDDGVQNPRAEDLVGKSLVYSATVILHSGSDMVQASRGIPITVSPYQHFT	358	
Db	300	DAVLTARKVLMGVPSNADALVGKSLVYSATVILHSGSDMVQASRGIPITVSPYQHFT	359	
QY	359	KTPKFKGMPFDMVFTNPDGSPAPRVAVAGEBTVQSLTQDGVAKLSINTHPSQK	418	
Db	360	KTPKFKPAMPFDMVFTNPDGSPASKVLVVTQGSN-AKALTODDGVAKLSINTHPSRQ	418	
QY	419	PLSITVTRTKQELSEAOATRMQALPYSTVGNNSNVLHLSVLTRELPGETLMNVNELLR	478	
Db	419	PLTIVTRTKQELSEAOATRMQALPYSTVGNNSNVLHLSVLTRELPGETLMNVNELLR	478	
QY	479	MDRAHEAKIRYTYLIMNKGELLKAGQVREPQGLVLPISITDTPSPRLVAYYTLI	538	
Db	479	TPDGEAKIRYTYLIMNKGELLKAGQVREPQGLVLPISITDTPSPRLVAYYTLI	538	
QY	539	GASGQREVVDVSVVDVSKVSGSLVKSQSDRQPVPGQMTLKIEGDHAGRWLVAV	598	
Db	539	GASGQREVVDVSVVDVSKVSGSLVKSQSDRQPVPGQMTLKIEGDHAGRWLVAV	597	
QY	599	DGKGFVLMNKNKLTOSKIWDVVEKADICTPGSGKDYAGVPSDAGLFTSSGQQTQRA	658	
Db	598	DGKGFVLMNKNKLTOSKIWDVVEKADICTPGSGKDYAGVPSDAGLFTSSGQQTQRA	657	
QY	659	ELQCQPPAARRRRSVQLETKMDKVGKY-PKELRCCEDGMRENPMRPFSCORRTFISLG	717	
Db	658	DLECTPAARRRRSVQLETKMDKAGQVTDKGLRCCEDGMRDIFPMYSCORRRLITQG	717	
QY	718	EACKKVFLDCCNYITELRRQHARASHGLARSLNDDEDIABENIVSRSEFPESMLNVED	777	
Db	718	ENCIRAFIDCCNHIKTLREQHRRDVLGLARSELEEDIIPEDIISRSHFPQSLWLTIEE	777	
QY	778	LKEPPKNGI-STKLNNIFLKDSTTTWEILAVSMDSKGI-CVADPEVTVMQOFFFDLRLPY	837	
Db	778	LKEPEKNGISTKMMIFLKDSTTTWEILAVSLDKGI-CVADPEVIRVMQOFFFDLRLPY	837	
QY	838	SVVRNEQVEIRAVLYNRQNELKVRVELLHNPAFCSLATTRHQQQITITIPPKSSLSVP	897	
Db	838	SVVRNEQVEIRAVLYNREQEELKVRVELLHNPAFCSMATKNRYFQTIKIPPKSSVAVP	897	









QY 1335 EGKGGTSLVVTYHA--KAKQOLTCNKEDLVKTIKPAETEKRPQDAKNTMILEICTRY 1392  
 Db 1291 SGNGEATLSVVTLYALPEKDS-DCESFSLVLTWMDKISH--EDAKESFMTIEVLY 1347  
 QY 1393 R-GQDQATMSILDISMNTGAPDPTDLKQLANGVDRIYSKYELDKAFSDRNLTLLIYLDKV 1451  
 Db 1348 KNSERDATMSLTDIGLITGTVTDNLQNLKGRERYIEKFMOKVLSRGLIYLDKV 1407  
 QY 1452 SHSDDDCLAFKVOHYFNVELIQGAVKVVAYYNLESCTRFVHPEKEDGKLNKCRDEL 1511  
 Db 1408 SHKLEDRISKIHVRQVGVLPAAVSVVEYIN-QRCVKFHPHQEGGTLRSLCLGDVC 1466  
 QY 1512 RCABENFIOKSDDKVTLBERLAKACEPGVDVYVYKTRLVKQLSNDPFDYIMAIETQTKS 1571  
 Db 1467 TCABESCMOKGEPDV--QRIDKACAGLDVYKATVVDKLTITDVTYTKIDLVIXP 1524  
 QY 1572 GSDESVQVQORTFISPIKREALKLEKHYLMWGLSSDFWGEKP-----NLSYIIIGKDTW 1627  
 Db 1525 GTDEGVGKRDPMFLAYCREALGLMGQKTYMIMGKSEDLHRVDEKGLLOYKYVLGEQTW 1584  
 QY 1628 VEHWPEDDECODEENKQCODLGAFTESMVVFGCP 1662  
 Db 1595 IEYWPQOECTRDYREVCLGIDEFINQITFGCP 1619  
 RESULT 9  
 A27602  
 Complement C3 - rabbit (fragment)  
 A:Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subunit;  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 15-Dec-1988 #sequence\_revision 07-Oct-1994 #text\_change 09-Jul-2004  
 C:Accession: A27602  
 R:Kusano, M.; Choi, N.H.; Tomita, M.; Yamamoto, K.; Migita, S.; Sekiya, T.; Nishimura, S.  
 Immunol. Invest. 15, 365-378, 1986  
 A:Title: Nucleotide sequence of cDNA and derived amino acid sequence of rabbit complement  
 A:Reference number: A27602; MUID:87006907; PMID:3019881  
 A:Accession: A27602  
 A:Molecule type: mRNA  
 A:Residues: 1-726 <KUS>  
 A:Cross-references: UNIPROT:P12247; GB:M32434; NID:G164862; PIDN:AAA31190.1; PID:G164863  
 C:Comment: Complement C3 contains two chains, formed by removal of four residues and lin  
 alternative-complement pathways, releases the C3a anaphylatoxin from the amino end of  
 native-complement-pathway C3/C5 convertase.  
 C:Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.  
 C:Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign pa  
 e classical-complement-pathway C3/C5 convertase. The activity of C3b is regulated by pro  
 C:Comment: The major site of synthesis of this plasma protein is the liver.  
 C:Superfamily: alpha-2-macroglobulin  
 C:Keywords: acute phase; complement alternate pathway; complement pathway; glycoprotein;

Query Match 35.8%; Score 3082; DB 2; Length 726;  
 Best Local Similarity 79.6%; Pred. No. 8.3e-181;  
 Matches 578; Conservative 74; Mismatches 74; Indels 0; Gaps 0;  
 QY 938 MNKTAVVRLDPERLGRGVQKEDIPPADLSQVDPDTSETRILLOQTPVAQMTEDAVDA 997  
 Db 1 MNKTAVVRLDPERLGRGVQKEDIPPADLSQVDPDTSETRILLOQTPVAQMTEDAVDA 60  
 QY 998 ERLKHLVITSGGCEQNMWMTVTIAVHYLDTEQWKEGLEKRGQALELIKGYTOOL 1057  
 Db 61 ERLKHLVITSGGCEQNMWMTVTIAVHYLDTEQWKEGLEKRGQALELIKGYTOOL 120  
 QY 1058 AFRQPSAFAFVKRAPSTWLTAYVVKVFLAVNLIAIDSQVLGAVKWLILEKQPDGV 1117  
 Db 121 AFKQPSAYAAFLNRPSTWLTAYVVKVFLAVNLIAIDSQVLGAVKWLILEKQPDGV 180  
 QY 1118 FQEDAPVTHQMICGLNNKEDWALTFVLISLOAKOICEBOVNSLPGSITKAGDELE 1177  
 Db 181 FQEDAPVTHQMICGLNNKEDWALTFVLISLOAKOICEBOVNSLPGSITKAGDELE 240  
 QY 1178 ANYMNLQSTVTAITAGYALQMGRLKGLPLNKELTTAKDKRWEDPGKOLYVNEATSYAL 1237

Db 241 ANTMNLQRPYSVAIAAYAWAQDQLRGLAFNLKFLSKAKENRWEPPQORLYNVEASSVAL 300  
 QY 1238 LALLQLKDFDFPPVVRVWNEQRYVGGYGSTQATFMVFOALAQYOKDAPHOBLNDVS 1297  
 Db 301 LALLURLDPSFPVVRVWNEQRYVGGYGSTQATFMVFOALAQYQTDVDPHKKLNWVS 360  
 QY 1298 LQLPSRSSKITHRIHWESASLLRSEETKENEGFTVTAEGKQGGTSLVVVTHAKAKDQLT 1357  
 Db 361 IQLPSRSSPVKHRIVWDSASLLRSEETKENOGFSLTAQKGGTSLVVVTHFAKVGKVT 420  
 QY 1358 CNKFDLVITIKAPETEKPEQDAKNTMILEICTRVRGQDQATMSLTDISMTGTGAPDPTDD 1417  
 Db 421 CKXFDLRVNIKTAPEFVKKAPQDAKNTMILGHCTRILGDEDATMSLTDISMTGTGAPDPTDD 480  
 QY 1418 LKLANGVDRIYSKYELDKAFSDRNLTLLIYLDKVSHESSDCLAFKVOHYFNVELIQGAV 1477  
 Db 481 LNLSTGVDRYISKVELNKAFAFNKNTLLIYLDKISHSREELAFKVOHYFNVELIQGAV 540  
 QY 1478 KVVAYYNLESCTRFVHPEKEDGKLNKCRBELCRCAEENCFIOKSDDKVTLBERLDCAC 1537  
 Db 541 KVVAYYNLESCTRFVHPEKEDGKLNKCRBELCRCAEENCFIOKSDDKVTLBERLDCAC 600  
 QY 1538 EPGVDYVYKTRLVKQLSNDPFDYIMAIETQTKSGSDEVQVQORTFISPIKREALKLE 1597  
 Db 601 EPGVDYVYKTRLVKQLSNDPFDYIMAIETQTKSGSDEVQVQORTFISPIKREALKLE 660  
 QY 1598 EKXHYLMWGLSSDFWGEKPNLSYIIIGKDTWVEHWPEDDECODEENKQCODLGAFTESMV 1657  
 Db 661 DGXHYLMWGLSSDFWGEKPNLSYIIIGKDTWVEHWPEDDECODEENKQCODLGAFTESMV 720  
 QY 1658 VFGCPN 1663  
 Db 721 VFGCPN 726  
 RESULT 10  
 150806  
 Complement component C3 - Japanese lamprey (fragment)  
 C:Species: Lampetra japonica (Japanese lamprey)  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
 C:Accession: I50806  
 R:Nonaka, M.; Takahashi, M.  
 J. Immunol. 148, 3290-3295, 1992  
 A:Title: Complete complementary DNA sequence of the third component of complement of la  
 A:Reference number: I50806; MUID:92251197; PMID:1578150  
 A:Accession: I50806  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1673 <NON>  
 A:Cross-references: UNIPROT:Q00685; GB:D10087; NID:G222887; PIDN:BA000983.1; PID:G222888  
 C:Superfamily: alpha-2-macroglobulin  
 Query Match 28.9%; Score 2490.5; DB 2; Length 1673;  
 Best Local Similarity 33.9%; Pred. No. 6.2e-144;  
 Matches 587; Conservative 353; Mismatches 653; Indels 137; Gaps 41;  
 QY 10 LLLLTHLPLALGSPMYSIITPNILRSEETMVLAEHDAQGDVPVTVTHDFPGKKLV 69  
 Db 1 VLLMSVGVSVTQDPWVLSVPSVILIGSDVNVVLDHAASTEDSVVVRASEFLTQKQLA 60  
 QY 70 SSKTV--LTP--ATNHMG----NVTFTIPANREFK-----SEKGENKVTVQATFGT 114  
 Db 61 TQTTITQLDPAIATLKLGFDIENPKTNSASTGHVRLVAKVESKSNKEITAH----- 116  
 QY 115 QVVEKVLVLSQSGVLIQTDKTIYTPGSTVLYRFTVNHK---LLPVGRVWVMIENPE 171  
 Db 117 -----LLSYRSHGVVQTDKPIYTPDKVKYRFPNVRDHRIPVRQSMVTDVNNAD 169  
 QY 172 GIPVKD--SLSQNLGVLPLSWIDPELVNNGQKIRAYYENSPOQVSTFEFEVKEVL 229  
 Db 170 GVTVRQIKTIKATDGIQVDTGTSFTIPAIKSHGTWIKIFARMSGAENINSSAEFDVREYL 229  
 QY 230 PSEFVIEPTEKFPYIYNEKGLVETITATFARFLYKGVKVEGTAFVIFGIQDGEORISLPESL- 288



Db 230 PTEVKINPQRFVHI-NDEEFVWDITANFYFNQELVSGTAVRYFLENGD---VPLVD 284  
 QY -KRIPIEDSGEVVLSRKVLQVQNPRAEDLVGSLVSVATVILHSGSDMWQAEBSGIP 347  
 Db 285 SSSTTVAGEGSLKKEKLLK-LFFNAKOLLAFSLTKITVJUSSQAETEEAEVLGK 342  
 QY IVTSPYQIHTFKTPKFKGMPDFLVMFYTNPDGSPAYRVPVAVQ---GSDTV---OSLT 401  
 Db 343 IVERSQITATKTSRYFKPELPEFYIQVEVRNADGSPSKVEDVWAKVQVGSATINPQKMT 402  
 QY QDGVAKLSINTHPSKXPSITVRKQKQELSAEQATMTQALPYSTVGSNNVHLISVL 461  
 Db 403 DSNGLTSFTV-TPPNVQNTVVRIDERHPSN-EGELVYTAQKVA---SASTVHIDVT 456  
 QY RTELPGETLNVNLLRMDRAHEAKIRYVYTLIMNKLKAGRGVREPODLVPLSLI 521  
 Db 457 RI-MRLGETLNVFLTAKTQLN--AVHTFTYMLVTRGVIVKTRKTESGGGPNVRPI 513  
 QY TTDPIFSELVAVYTLIGASGOREVVADSVVMDKSCVG--SLVVKSGGSEDRQPVPGQ 579  
 Db 514 TPDMAFRFLAYIILPGG---EIVADSVTVETELCKSQVSLKGRPTLE---PKA 565  
 QY QMTLKIEGHDGARVLVAVDKGVFLVANKKNTQSKINWDVVEKADICTPGSGKDYAGVF 639  
 Db 566 MLTDLIGSPDARVGLVAVQVAVNKRHLTQDRVWKAMETFTDGTCTABGGAGRPGVF 625  
 QY SDAGLTFTSSGGQTAQRAELQCPQAPARRRSVOLTEKMDKVKYKPKELKCCEDGMR 699  
 Db 626 SDAGLALITSKGLNTDRSEIGCPKVPKRPQLSMLQIRRE-AEKTOFRKCCVDGLK 684  
 QY ENPMRFSQRRTRFISLGBACKKVLDCNCTITELRQRAHSLGLARSNLDF-DIAE 759  
 Db 685 MSPTGQCCERLKRVTGPKCEVDAFLQCCKAAEYRKESLCAKTFLRRNDFMELDLNNE 744  
 QY ENTIVSEPESELWNVDELKPPKNGISTKLMTFLKDSITTWELLAVNSDKKGLCVA 818  
 Db 745 DEVMNMAFPQSGWGN--KYKNSCKYGRHPQI-RLQLPDTITWNNQVVISKTRGVCLA 801  
 QY DPEVTWQDFFIDLRLPVSVRNEQVEIRAVLYNQRQBELKVRVLLHNPAFCSUAT 878  
 Db 802 DPLLVSTKDFFKLHLPVSVRGEGTEIRVILVNYME-ESLITLITEMDIVESICSTKS 860  
 QY KRHOQTIITPPKSSLSVPVTVPELKTGLQVEVKAAYTHFISDGVKRSKVVPEGIR- 937  
 Db 861 GAPSQKSVKGAWVSPFVPLKIGBHHSIRSVYGRFTGQGVQKILRVAPEGVVD 920  
 QY -MKNVAVRTLDERLREGVOKEDIPPAADLSQVDPDTESETRILLQSTPVAQMTEDAVD 996  
 Db 921 IRSERSVHVERETP---FINETISP---DVVENSDELVTFSVKGDELAETWVNCGL 972  
 QY AERLKLHIVTPSCGQBNMIGTPTVIAVHYLDETEQWEKPGLEKRCQALELIKGYTQ 1056  
 Db 973 AKSISNLIOIPTGCGEONIKWAPTLLIYLDVSOEWKEGLHRREAEATGFLAQGSRE 1032  
 QY LAFRQSSAPAFKPAKPTWLTAVYVVFSLVAVNLIAIDSVTCGAVKMLILEKQKPDG 1116  
 Db 1033 LSYRKADHSYAAPIKRPSTWLTAFVVKVYSLAKEVIVDQELCGPVEWITIKRQNSDG 1092  
 QY VFQEDAPVHQEMIGELRNNEKDMALTAFLVLSQEAKDICEBQVNSLPQSTIKAGDGL 1176  
 Db 1093 SYREDGPVIREMCGGV-GGTEGHVSMATFILIIGQQAQEGCVSVNPKSNMRAVQFL 1151  
 QY BANYMNLQSYTVAINAGYALAOQGRKGP-----LLNKLFTTAKDKNRW---EDP 1223  
 Db 1152 ASKVSLLKRYMTIATRYALA---LQDPESEAAHSWKKLENTRTFESKGRHYWKAET 1207  
 QY 1224 GKOL---YVNEATS YALLALQLKDFDPVPVFWNLNEQRYGGYGGYSTOATFWPQAL 1279  
 Db 1208 SHVLRVSAISVEATAGLLTYLRKKDYESAREIVDMLTEQRNYGGGQSTDTILALQAM 1267  
 QY 1280 AQYQKADPHQELNLDVSLQFSRSKITHRLHWSASLLRSBETKENEGFTVTAEGKG 1339

## RESULT 11

T30517  
 Complement C3-Q2 - common carp  
 C:Species: Cyprinus carpio (common carp)  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #ext\_change 09-Jul-2004  
 C:Accession: T30517  
 R:Nakao, M.; Obo, R.; Mutsuro, J.; Fujiki, K.; Nonaka, M.; Yano, T.  
 submitted to the EMBL Data Library, July 1998  
 A:Description: Sequence diversity of the third complement component (C3) from a bony fish  
 A:Reference number: 220840  
 A:Accession: T30517  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1230 <NAK>  
 A:Cross-references: UNIPROT:Q9YIA5; EMBL:AB016215; NID:d123045; PID:d1037610; PIDN:BAAS

Query Match 28.3%; Score 2438.5; DB 2; Length 1230;  
 Best Local Similarity 40.58; Pred No 5, 7e-141;  
 Matches 522; Conservative 231; Mismatches 406; Indels 129; Gaps 19;

QY 388 QSLTQGGGVAKLSINTHPSQ-----KPLSIT-----VTKKQELSEAEQATR 439  
 Db 46 QDYSREDLVNKLIVYKHPKXMDILSKSVTLTAANCFOILTDFKATKDPEDLDEQQAVR 105  
 QY 440 TMOALPYSTVNSNNVLSLVRLTELPGETLNVNLLRMDRAHEAKIRYVYTLIMNKR 499  
 Db 106 KMTAQAEVTKGSSQNYLHIGIDAELQIGDSVRV--YLTGQIPGVKQDQFYMLSKQG 163  
 QY 500 LLKAGRVREPGQDLVPLSITTDIPSPRLVAYTLIGASGOREVADSVWDVSKSC 559  
 Db 164 IVKVDREKRG-QGSLVTLPTVTVKDMVPSFRFVAYTHV-----GSSEWSDSVVWDVQDTC 218  
 QY 560 VGLSVKSGGSEDRQPVPGQMTLKEGHDGARVWLVAVDKGVFLNKKNKLTQSKIWV 619  
 Db 219 MGTLIQVKOKLNIYDT-GDEVKLQITGDPGKVLVWVDKAV---NKNRLTQIQIMDV 273  
 QY 620 VEKADIGCTPGSGKDYAGVPSDAGLTFTSSGGQTAQRAELQCPQAPARRRSVQLTEKR 679  
 Db 274 IEKHIDGCTAGGGRDSMGVFTDAGLMPFESNTAGTNTTRTPDCPFLSKRRRS----- 326  
 QY 680 MDKGVKVPKELRKCCEDGMRPNRPFSCQRRTRFISLGBACKKVFLLDCNCTITELRRGHA 739  
 Db 327 -----DDDDYVD----- 334  
 QY 740 RASHLGLARSLNLDDEDTAENIVSRSEFPESWLNWVNDLKEPPKNGISLKLNNILKDSI 799

Db 335 -----SEIVSHQTQFRESNWEIEDLCKDCSTPATEKV--IYLKDSI 374  
Qy 800 TTWEILAVSMDKKGICVADPFVETVMQDFDRLRPLYSVVRNEQVEIRAVLYNRQOE 859  
Db 375 TTWQILAVSLPTLIGICVAEPBEMVFKLFDLKMPSAVRGEQLEIKAIHNYTPNKQ 434  
Qy 860 LKVRVELLHPAPCSLATTERRHQOITIPPKSSLSVPYIVPLKTLGCEVEKAAVYHH 919  
Db 435 -KVRVEPMEWEDVCSFKGKRTTIVSEKSSISVSIVIPMLGNHMEVKAASADA 493  
Qy 920 FIDGVRKSLKVVPEGRMKNKTVAVRTLPERLREGVQKEDIPPADLSDQVDPDESBR 979  
Db 494 IYTDGVRKQLKVVSEGLTSHVGRKVELNPVKNQKPLVFKSVIPA---DRLDPTADMY 550  
Qy 980 ILLOGTPTVAQWTDVDAERLKLIVTBSGCEONMIGTPTVIAVHYLDTEBQWKEGL 1039  
Db 551 IAITGSEITQVBOA:SGSFMGRLIVQPSGSGEONMMLTFLIATHYLDSTSQWDTVM 610  
Qy 1040 EKRGALJELIKGYTOQLAPRPSAPAFVVRAPSTWLTAYVVKVSLAVNLIAIDSOV 1099  
Db 611 ERNEAVNYINTGYRQLGYRKSDGSGYAAWTHRPSSTWLTAYVAKVFSMANNIVNEEV 670  
Qy 1100 LCGAVKLLILEKOPDGFCEADPVTHQEMIGGLRNNEKDMALTAFLVLSIORAKDICE 1159  
Db 671 ICSALKWLLIHKQTLGSKFEDSAVLOGEMVGIQ-SKDRDASLTAFFVIAWQEAETICA 729  
Qy 1160 EQVNSLPGSTKAGDFLEANYMNLORSYVYIAGYALAOQWGLKGLPLTKLTTAKORR 1219  
Db 730 GSVASLHESIRKAVSFLEGRPLQPTNPYAVAMMSYAMANEKLNKILMKHSSQEAQRS 789  
Qy 1220 WEDPGKOLYNVEATSALLALLOLQDPFVPPVVRWLNQRYGGYSGSTQATNMVFOAL 1279  
Db 790 WTVPGQHHLSEATAYAVLALVAKAFDRAGEAVHWNQOHHYGGSTQATINLVFOAV 849  
Qy 1280 AQYQKADPADHQLNVLVSLOLPSRSKITHRIHWEASALLRSEETKENEGFTVTAEGKQ 1339  
Db 850 AEYRTOVKDOQNFNLDELVSAGRGKPVRYTIKRENAHLTWGKMEINKEFNVTARGYGR 909  
Qy 1340 GTLSVWTHYHAKADQLT-CNKFDLKVITIKPAPETEKRPQDAKNTWI---LEICTRYEG 1394  
Db 910 ATLSVLTLYARPVEKSKDCTFDLTKM-----EKNBAKHGTIASYKLTMDFIYS 962  
Qy 1395 DQ-DATMSILDI SMGTGAPDPTDQLKQANGVDRIYSKYVELDKAFSDRNTLIYLDKVS 1453  
Db 963 DKTATMTILDVGLPTGFEVESDLKQLSGKERYIQKPEMKNVLSERGSJLYLDKVS 1022  
Qy 1454 SEDDLAFKHYQVFNVELIOPGAVKYVAVYNLEESCTRYHPEKSDGKLNKLCRDELQRC 1513  
Db 1023 KEKVVSFRHQY:YVUGLLQPAATYIYSPDARCTMFIHERTDAIYRLCKGDLCLC 1082  
Qy 1514 AEENCFIQSDDKVTLEERLDKACEPGVDYVYKTRLVKVLQSLNDFDEYMAIEQTIKSGS 1573  
Db 1083 AEENCSFKKNN-VROEERLNRVCEPGMDYIYKVTVVGMDLKQDSDIYDMKVELVKEGT 1141  
Qy 1574 DEVQGOORTFISPIKREALKLEEKHYLMWGLSSDFWGEKENLSYIIGKDTWVHWPE 1633  
Db 1142 DEDAGKVRQFLAPSCREHLGLVECKSYLNGRSDDLPELGSLOYVFGQTVWEIYPT 1201  
Qy 1634 EDEQDEENQKQCODLGAFTESMVVFGC 1661  
Db 1202 REESQTRHRRVIGISELQNSL-RYGC 1228

RESULT 12  
S21045  
complement protein homolog - inshore hagfish (fragment)  
C:Species: Eptatretus burgeri (inshore hagfish)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
C:Accession: S21045  
R:Jishiguro, H.; Kobayashi, K.; Suzuki, M.; Titani, K.; Tomonaga, S.; Kurosawa, Y.  
EMBO J. 11, 829-837, 1992  
A:Title: Isolation of a hagfish gene that encodes a complement component.  
A:Reference number: S21045; MUID:92192016; PMID:1372251

A;Accession: S21045

A;Molecule type: mRNA

A;Residues: 1-1620 &lt;ISH&gt;

A;Cross-references: UNIPROT:P98094; EMBL:Z11595; NID:G62774; PIDN:CAA7677.1; PID:G62777

C;Superfamily: alpha-2-macroglobulin

C;Keywords: plasma

Query Match 25.0%; Score 2148; DB 2; Length 1620;

Best Local Similarity 31.7%; Pred. No. 5.9e-123;

Matches 535; Conservative 338; Mismatches 69; Indels 124; Gaps 44;

Qy 13 LTLPLALGSPWYSIIIPNIRLESEETMVLAEADAOQDVEVTVTHDFFPKKLVSSE 72

Db 1 VLVIAFAAATSS--YDDLAVAILMVDQKK--ITEVH-----VLVNHPTGATL---DE 45

Qy 73 KTVLTPTATHMGNVTFITPANREFSEKGRNKFVTVQATF-CTQVVEKVLVLSQSYLF 131

Db 46 KRVKLGWDNKF--IAFTKLQVTPKEVEKEDFVRLMWKDGQGHMELIDPLTSRRGLVF 103

Qy 132 IQTDKTIYTPGSTVLYRIFTVNHKLLPVGRITVMNIENPEGIPV---KQDSSLSSQNLGV 188

Db 104 AQTDQPIYTPNDVNIRLFPVTRQLNPLSLVAVDMMNPDGVVVDRIEKNAFEVKWML 163

Qy 189 LPLSWDIPELVMGQWKIRAYYENSPOQVFSTETPEVKYVLPSEVIVEPTEKYYIYNE 248

Db 164 RP--FHVPAITSLGDMKIVSMWKDPQFNYSFGFKVEEYVLTPTDVSITSSQPYLHY-D 220

Qy 249 KGLEVITITARFLYKKGVEGTAFFVIFGIQDGEQRISLPESLKRIPIDGSGEVLRSKVL 308

Db 221 KAPTTHIKAMHIYKGVPMGRVAVRYGVKHQSKRTLLSTSSALARPEQGEAMHTLRQKHL 280

Qy 309 DGQVNPRAEDLVKSLYSYATVILHSGSDMVQABERSGPIVTSYQIHFTKTPKYEFGM 368

Db 281 EQVPDPLK--LIGQSUYVEASVISSDAEIEINSILDDIPIVASPYISKSKWTVFFFPKV 338

Qy 369 PDLMAVFTNPDGSPAYRVPVAVQ-----GEDTVQSLTQGDGVAKLSINTHSQKPLS 421

Db 339 FVIYKVLVLPDGPSPASGVPIKVSFSDSSGNWITQKRTKTDNGIAMQTIATARNKKLN 398

Qy 422 ITVTRTKQBLSEABQATRTMQALPYSTVGNNSNYLHLSVLTLELPGSETLVNVLMDR 481

Db 399 IKVQTEDELEQSQAEASFTIASYSS--PSGFIHLNAHREVKSPEG--HIVDFVFIKS 454

Qy 482 AHEAKIRYTYILIMKGRL---LKAGRVPRFGQDLVVLPLSITTDFTIPSPRLVAYTIL 538

Db 455 AAKDHVLFHNYLWISNGKIHNFLOSGRK----GDTTSVSLLTLPVLVQFRLVAFIL- 508

Qy 539 GASQREVVDVSWVDVVDKSCVG--SLVYKSGQSEDRQPVFGQMTLKIEGDHGRVVLV 596

Db 509 -PSG--ELVADSIIDVXDSCHAKLSLDVAGK---RLFSPRDNVNFDSGESDSWAVG 562

Qy 597 AVDKGVFLVNLKKNLTQSKINDVVEKADIGCTPGSGKDYGAVFSDAGLTFTSS--SQQT 654

Db 563 VVDXAAVYLDKKNLTANKVYKAMEASDLGCSVSGSGTKGPLVFRDAGLAIWAKISGWD 622

Qy 655 AQBAELQCPQAPARRRSVQLTEKMDKVYKPKELRCCEDGMREPMRSCORRTFI 714

Db 623 VK--DPCNGNTHRRKRLVL--EIAIERKASTYPAELRCCRDAAIESPLRUSCEERTKI 679

Qy 715 -SLGEACKKVFLLCCNYITELRQARASHILGLARSLNLDIIAENIVSRSEFPESMLW 773

Db 680 HDEGECQSTFLECCCKHVEEELLIAWEEBDDGLRSGQEDFMIQESQVIRSHFESFMW 739

Qy 774 NVEDLKEPPNGIS--TKLMNIFLKSITTWELLAVSMDKKICVADPPFVTVNQDFFI 831

Db 740 EIKUSRAENKSRITKKN-----PDSITTWIDIQAVEVSQSGKLCVGSLSLTVFKQFFL 795

Qy 832 DLRPLYSVVRNEQVEIRAVLYNRQONELKVRVELLHNPAFCSLATTKRHQQTITPPK 891

Db 796 KVHTPYALKQYEQVELRVYIYNY--NNQVKGIEIQVKGDGICTDAEQNEPLKSRFAVEKN 854

Qy 892 SSSLSPVYVPLKTLGQVEVKAAYVHHFISDGVKSLKVVPEG--IRMKNTAVRILDP 949

Db 855 SATSFEMVPLSSSDSSVSVLAVFGSDVDHDAVEKDLRVMPGNYEMSRSHSVQ--P 911  
 Qy 950 ERLGREGVQKEDIPIADLSQDVPDTESETRILLQTPVAQMTEDADAERLKLHIVTPSG 1009  
 Db 912 RRRG--GQGVIVVDNETPQNVVPGTEMGAFLSAQGNLVAETIQNTLKGSKISNLLRPRG 969  
 Qy 1010 CGSONMIGTPTVIAVHYLDETEOKKEG--LEKROGALELIKKGTYQQLAFQSSAF 1066  
 Db 970 CGSQNMVYTSITVMVARYLNRSDQWKNQDQPOLKRR--SFDITSGFASQUTRKPDYSY 1027  
 Qy 1067 AAFVKEAPSTLTAYVYKVFSLAVNLIAIDQVLCGAYKWLILSKQKPDGVPQSDAPVIH 1126  
 Db 1028 AAMLHRSSTLTATFAKVFQSQARQLVPIPVSEICGSVRML-MRKQDKDSFLESKEFVH 1086  
 Qy 1127 QEMIGLNRNNEKOWALTAFLVLSIQEAKDICEQVNSLPGSITKAGDFLEANYMNLQRS 1186  
 Db 1087 LNMGGVOTK---VLTISFVFIALLARESCINEVEGTVVVKAGHYLTSQANGLD 1142  
 Qy 1187 YTVAIAGYALAQGRLLKGLNKLFTTAKDN-----RWEDPGQLYXNVATSYALLA 1239  
 Db 1143 FPLAITAYALS-LWKVSDGAAKVTMTHTLTKSTGLQTEELIHWGSKNGKAAAVESTAYGLA 1201  
 Qy 1240 LIQLKDFDFVPVFWLNRQRYGGYGSTQATFMVFOALAQYOKDAPDHCENLNDVSLQ 1299  
 Db 1202 AICHESGEIETATWLSOSATFGYFQSTQDWTWALQALTFPSCQSRMKNDLSFKIR 1261  
 Qy 1300 LP-----SRSSKITHRIHWESALLRSBETKENEGFTVTSAGKQGTLSVVVWYHAKAD 1354  
 Db 1262 AEENGVDKEFQITN---DNAFQVQKPKVPVHGQITVTSAGTGQGITTFVKYREKVI 1317  
 Qy 1355 QLTCKNFDFLVTIKPAPETEKEPOAKNTMILEI---CTRYRGDQDATMSILDISMMTG 1410  
 Db 1318 KDCQKFSLEITNLNDQVQRROSINPEFNVRFICFRYLNRQEPGMVMDISLPTG 1377  
 Qy 1411 FAPDPTDLQLANGVDRIYSKVELDKAFSDRNTLIYLDKVSHEDDCLAFKHQYQNVNE 1470  
 Db 1378 FEAKKXDLDMKNLVDNYVQVEIRP-----GRVFLYLDKVNKDKCNVCFRLNQVPEIN 1432  
 Qy 1471 LIOPGAVKVYAYNLESTCTRFYHPKEKDGKLNKLCRDELCKCAEENCF-IKSDDDKVTL 1529  
 Db 1433 LVLPVTATFVEYEPDFRCSKYHFKMVEWP-DASCHGNI CNCLQRHCVELKGMADERN 1491  
 Qy 1530 EERLDKACBEGVDVYKTRLVQVLSNDFDEYVMAIETIKSGSDE-VQVQQRFTFISPI 1588  
 Db 1492 ADRNGNACR--AEYVFIIGVTKVTASYNINAAKTLVKKGMQDAINVGARRSFVPM 1549  
 Qy 1589 KCREALKLEKKHLMWGLSSDFWGEKPNLSYIIIGKDTVWEHWPED-----ECOD 1639  
 Db 1550 HCGKLNLVPGDIYLYVGMENAHWNSDRTQVLTSDTWFEKFPLESVCRILPSPASCQV 1609  
 Qy 1640 EENQKOC 1646  
 Db 1610 SENFKGC 1616

RESULT 13  
 T30885  
 Complement component C3 homolog - African clawed frog  
 C;Species: Xenopus laevis (African clawed frog)  
 C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
 C;Accession: T30885  
 R;Mo, R.; Kato, Y.; Nonaka, M.; Nakayama, K.; Takahashi, M.  
 Immunogenetics 43, 360-369, 1996  
 A;Title: Fourth component of Xenopus laevis complement: cDNA cloning and linkage analysis  
 A;Reference number: 220919; MUID:96186527; PMID:8606056  
 A;Accession: T30885  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-1683 <MOR>  
 C;Cross-references: UNIPROT:Q91741; EMBL:D78003; NID:g1183856; PIDN:BAA11188.1; PID:g118  
 C;Superfamily: alpha-2-macroglobulin

Query Match 24.6%; Score 2122; DB 2; Length 1683;

Best Local Similarity 31.5%; Pred. No. 2.5e-121;  
 Matches 554; Conservative 328; Mismatches 704; Indels 170; Gaps 49;

Qy 1 MGTPSGSLLLLL---THLPALGSPMYSIITPNILRLESEETMVLBAHQDQVPVTV 57  
 Db 1 MGP---PPLLYALVMAAVCMVAKAQPNFLVAPRIILHVGVKEVIGICIPWEPQGSPLPAG 57  
 Qy 58 TVHDFPQKLVLS-----SEKTVLTPATNMGVNTFTIPANREFKSEK----- 100  
 Db 58 DI-----KVELSURNQLSMIGSGSEKFLNKRND-----YTILNMLTSLRLICK 105  
 Qy 101 -----GRNKFVTVQA-TFGTVVEKVLVLSQSGYLFIQTDKTIYTPGSTVLYIFFTV 153  
 Db 106 LNERRSKRYQLVVKSDVLGKQKWSIPVAYQGYLFIQTDKSIYTPKRETHIRSFPLD 165  
 Qy 154 HKLLPVGRTVMVAIENPEGIPVKQDSLSQNLGVLPLSMDIPELVNMGMKIRAYVENS 213  
 Db 166 HVLPRTBEQVLSVFNACQGFQVRKVKISKD--SVVADNLQIPDIPSTPGVWRISVHYTDA 223  
 Qy 214 PQOVSTFEFEKYLPSFEVIVETPEKFPYIYNEKLEVTITARFLYKKGVEGFAFVIF 273  
 Db 224 PETNFTAEFEKYPVLPNFEVKIIP-EVPYFLMTKDSFTFRVEARYVYGEHVAGVGNRV 282  
 Qy 274 GI--QDGEORISLPESIKRIPEDSGSVVLSRKVLLDGVQNPRAEDLVGSLVYSATVI 331  
 Db 283 GITAQNG-KRYMLAGLEKQTLIDGTTVTIRWADIKELIQDDMG-NLLGTSLSIPAAVSV 340  
 Qy 332 LHSGSDMVQABRGIPVITSPYQIHFTKTPKYPKQMPDLMVFTVNDPGSPAVPVPVAV 391  
 Db 341 EKASGVLEKEFTSVKFEVSSPYKLDLSKTRYPICGTPAQIIVVEVSHIDGSPAAGVSVL 400  
 Qy 392 QGEJT--VQSIQTQDGVAKLSINTHPSOKPLSTVTRTKQELSEAEQATRTMQALPYSTV 449  
 Db 401 SKSNAQVFEYKTDNGVVAFHVNTASDEKCDIRKVPANFSDSSETESIT---LLPYTS- 455  
 Qy 450 GSNNNYIHLVLFTELPGETLNVNLFLLNRMDRAHEAKIRYTYLLIMNKGRLKAGQVRE 509  
 Db 456 -KASSYFLVAVPNQVLDPGSSFKVT--LKAIISNKQVKEIYYMVLNKGQLSLDSISRT 512  
 Qy 510 PGQDLVPLSITTDFTPSRLVAYITLIGASQREVWADSVWVDVKQSCVGLSVKGGQ 569  
 Db 513 EVNEMLI---TVKPSMTPSFVIAYYL-----GSEIISNVWVDVADCEKLEHASK 564  
 Qy 570 SEDRQVPVQGMQLIKBGDHGARVILVADKGVFVNKNKNTQSKINDVVEKAIGCTP 629  
 Db 565 ---KILAPGALKLDVTEGTATVLSAVDTAVYILNSKNKLTPOKMFKAMNAYDLGCS 621  
 Qy 630 GSGKDYAGVFSDAGLFTSSSGQQTQARAELOCPQFAARRRRSVQLTEKRMKVGYK-PK 688  
 Db 622 GGGKDFINVTDLAGLAFVSSAGY--TQINELGC-RVHQKKKAI DFOALTQOKAYSITT 678  
 Qy 689 ELRKCCEDGMRENPMRFS--QQRTRFISLGEACKVFLDCNNYITELRQHA--RASHL 744  
 Db 679 ELQCCQHGMMLPFGKMSRVCTKRAARVP-DPTCRKAFLDCCCEYAEENLKQLTLEKRTQ 737  
 Qy 745 GLAPS---NLDEDI IAEENIVSRSPFESLWNVNEDLKEPPKNGISTIKLMMIFLKDSIT 800  
 Db 738 CFGITQNVGSDDEDFADESDIQIRSFPELWRTVTV---YNGLFSSA--VYMPDSIT 791  
 Qy 801 TWEILAVMSDKKIGICVADPFVTVQDFFIDRLPYSVVRNBEQVEIRAVLYNQNQEL 860  
 Db 792 TWEIQAIGMSREKGFCAEPLKVKVFKDPHIYLRVPYSVKRFEQMBELPILNY--NNKDL 850  
 Qy 861 KVRVELLHNFAFCSLATTKRHHQ-QTITIPKSSLSVPVYVIVPLTKTGLQEVKAAVYHH 919  
 Db 851 EVKYWMECAEDICSPSGSDSKPLIKVTVGANSALPFPVWVPIKSNPVSV-VALGRS 909  
 Qy 920 FISQGVKSKLVPEGEIRNMKNTVAVRTLPBRLGREGVQ-KEDIPPADLSQDVPTESET 978  
 Db 910 FVSDGVKAMKIVKEGASVFEESY-IIPADITRRSIDFDEBFP---SNMIPGDPRS 964  
 Qy 979 RILLQGTVPVQMTEDAVDAERLKLHIVTPSGCGEQNNIGMTPTVIAVHYLDETEOKKEG 1038

Db 965 SIKVTMDSSMNTINNSLGADGKSLIRVPYCAEQMTISTSPGVYALRYLDHTKWNLLS 1024  
 Qy 1039 LEKQGALEIKKGYTQOLAPROPSSAFARVABSTWLTAYVVKVPSLANVLAIDSQ 1098  
 Db 1025 PDRKDEGLNRRQGLILQFKADGSGYGAWLHRTSTWLTAFVVKVSLCNYIDVNV 1084  
 Qy 1099 VLGCAVKWLEKQKPGVFEQADPVHQIMIGGLRNNEKDMALFVLLISLOBAKDI 1158  
 Db 1085 DIRLSAQLV-ATMOKDYGAFQKSVIHQDMLGG-STTIDAESVITAYVTVSLYHSL 1142  
 Qy 1159 EEQVNSLPGSTITAGDFLEANYNMLORSYTVATAGYALQMGF--LKGPLNKELTAK 1215  
 Db 1143 SEDNVAVKSLKAVDRLKGLTIKHPYSALUTYVLTALTSKOSVLKDKAYNKLMSRAQ 1202  
 Qy 1216 -DKRWEDPGKLY-----NVEATSYALLALLQKDFDFVPPVVRWLNQRYYGGY 1266  
 Db 1203 GDNPK-----KELVFGPKGTALAVETTSYVLLITLRLGNKIDAEKMYTWLSQYNGG 1257  
 Qy 1267 GSTQATFMVQALAQY--QXADPHQELNDVLSQLPSRSSKITHRIHESASLRSBE 1324  
 Db 1258 KSTQDTWMALEALSEYWRIFKQDDNTLEIVNSLEKSHQSKFRLKEDNLOEELRSMGT 1317  
 Qy 1325 KENEGFTVTAEGKGGTSLVWVYHAKADQDLTKCNFKDLKVTIKPAPET----- 1373  
 Db 1318 K----FNKVSQKGGKILTVIKVYMLQMTQSTELGLDVTVSDACEAGNEDDDYD 1373  
 Qy 1374 EKRQDQAKNTWLEICTRYRGDDAT-----MSTLDISM 1407  
 Db 1374 DEEMADREPPIVHDLRRSRREAFTEKEVKLLYEVCLMKKSNVTLSGWALVDTL 1433  
 Qy 1408 MTGAPDDBLQLANGVDRYISKYELDKAFSDENTLIYLDKVSHEDDCLAPKHQYF 1467  
 Db 1434 LSGIEPIDDLKLKLAESSEYISHYEQP-----GRLLYFDKVPDWT-DVAFEAQTV 1487  
 Qy 1468 NVELIQGAVKVAAYVNLSESTFRFVHPEKEDKLNKLCRDELRCABENCFIOKSDKY 1527  
 Db 1488 KVSLLQASAVIDFYEPNFKCTVFGAPSKPNFVSTLCSGVQCAELCPKLSPPKS 1547  
 Qy 1528 TLEBRDLKAC-BPGVDVYVYKTRLVQLNSNDPEYIMAEQITKSGDE-VVGQOQRTPI 1585  
 Db 1548 TETERTQFACVSPRVFVGVYVMEVIAEQAFVYVTAIRILEVLQRNSDDIKPEPTRLFY 1607  
 Qy 1586 SPIKREALLKLEKHYLMGLSSDPFGKPNLSYIIGKDTWVEHWEPEDECOEENOKQ 1645  
 Db 1608 QPMCK--MRLAKETEYLINGQDQGVTKNEGHYILERKFWVEELPGEQKCAATRYNF 1665  
 Qy 1646 CODLGAFTSMVWFGC 1661  
 Db 1666 CTDARFNMDKYKNGC 1681

## RESULT 14

CSHU  
 Complement C5 precursor [validated] - human  
 N:Contains: C5a anaphylatoxin; C5b  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 09-Jul-2004  
 C:Accession: A40075; A27689; A01267; A01266; S15121  
 R:Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Hunt, A.; Wetsel, R.A.  
 J. Immunol. 145: 362-368, 1991  
 A:Title: Complete cDNA sequence of human complement pro-C5. Evidence of truncated transo  
 A:Reference number: A40075; MUID:91079575; PMID:1984448  
 A:Accession: A40075  
 A:Molecule type: mRNA  
 A:Residues: 1-1676 <HAV>  
 A:Note: 518-Ser was also found  
 R:Wetsel, R.A.; Lemons, R.S.; Le Beau, M.M.; Barnum, S.R.; Noack, D.; Tack, B.F.  
 Biochemistry 27, 1474-1482, 1988  
 A:Title: Molecular analysis of human complement component C5: localization of the struct  
 A:Reference number: A27689; MUID:98209511; PMID:3365401  
 A:Accession: A27689  
 A:Molecule type: mRNA

A:Residues: 412-1676 <WET>  
 A:Cross-references: GB:M65134; GB:M18879; NID:g179691; PIDN:AAA51856.1; PID:g179692  
 R:Parmander, H.N.; Hugi, T.E.  
 J. Biol. Chem. 253, 6955-6964, 1978  
 A:Title: Primary structural analysis of the polypeptide portion of human C5a anaphylato  
 A:Reference number: A01267; MUID:79005687; PMID:690134  
 A:Accession: A01267  
 A:Molecule type: protein  
 A:Residues: 678-751 <FER>  
 R:Lundwall, A.B.; Wetsel, R.A.; Kristensen, T.; Whitehead, A.S.; Woods, D.E.; Ogden, R.  
 J. Biol. Chem. 260, 2108-2112, 1985  
 A:Title: Isolation and sequence analysis of a cDNA clone encoding the fifth complement  
 A:Reference number: A01266; MUID:85130937; PMID:2579066  
 A:Accession: A01266  
 A:Molecule type: mRNA  
 A:Residues: 412-854; S1ALSPRLCEKNGISCHCKRLIPGSSDSPASQVAGITGTHHAQPT' <LUN>  
 A:Cross-references: GB:K02874  
 A:Note: the carboxyl-terminal part of the sequence in this report appears to be derive  
 R:Bohnack, J.F.; Mollison, K.W.; Boko, A.M.; Ashworth, J.C.; Hill, H.R.  
 Biochem. J. 273, 635-640, 1991  
 A:Title: Group B streptococci inactivate complement component C5a by enzymic cleavage a  
 A:Reference number: S15121; MUID:91144547; PMID:1996961  
 A:Contents: annotation  
 C:Comment: Complement C5 contains two disulfide-linked chains, formed by removal of fou  
 (beta and alpha' chains).  
 C:Comment: Activation of C5 initiates the spontaneous assembly of the late complement c  
 is the foundation upon which the membrane attack complex is assembled.  
 C:Comment: C5a has potent spasmogenic and chemotactic activity.  
 C:Genetics:  
 A:Gene: GDB:C5  
 A:Cross-references: GDB:119734; OMIM:120900  
 A:Map position: 9q34.1-9q34.1  
 C:Superfamily: alpha-2-macroglobulin  
 C:Keywords: complement alternate pathway; complement pathway; cytolysis; glycoprotein;  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-673, 678-1676/Product: complement C5 #status predicted <C5B>  
 F:19-673, 752-1676/Product: C5b #status predicted <C5B>  
 F:19-673/Product: complement C5 and C5b beta chain #status predicted <C5BB>  
 F:678-1676/Product: complement C5 alpha chain #status predicted <C5A>  
 F:678-751/Product: C5a anaphylatoxin #status experimental <C5T>  
 F:752-1676/Product: C5b alpha' chain #status predicted <C5BA>  
 F:567-810, 634-669, 698-724, 699-731, 711-732, 866-1527, 1101-1159, 1375-1505, 1405-1474, 1520-1  
 F:741/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:751-752/Cleavage site: Arg-Leu (C5 convertase) #status experimental  
 F:911, 1115, 1630/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 21.4%; Score 1840; DB 1; Length 1676;

Best Local Similarity 28.2%; Pred. No. 4.8e-104;

Matches 481; Conservative 343; Mismatches 759; Indels 124; Gaps 47;

Qy 26 YSIITPNILRSEETWLEAHDAQDVPVTVTVHDFPGKVLVLSSEKTVLPATNHMGN 85  
 Db 23 YVISAPKIFRVGASENIVIGVYEAFTATISIKSPDKKFSYSSGHVLSSENKFNQS 82  
 Qy 86 VTFTIPANREFSEKGRNKFVTVQATFGQVVKVLSQSLGVLFITDXTIYTPGSTV 145  
 Db 83 AILTIQP-KQLPGGQNPVSVYLVVSVVYVSVVYVSVVYVSVVYVSVVYVSVVYVSVV 141  
 Qy 146 LYRIFTVNHKLPVGTVMVNIENPEGIPVKQDSLSQNLGVLP-SWDIPELVNMGOW 204  
 Db 142 KVAVYSLNDLRPKARETVLTIDEGSEV--DMVEEDHIGITISFPDFKIPSPRYGMW 199  
 Qy 205 KIRAYENSQQQVSTFEFVKVYVLPSPFEVIVPEKPYIYNEKGLVTTAFLYKCK 264  
 Db 200 TIRAKYKEDFTTGTATYFVKVYVLPSPFEVIVPEKPYIYNEKGLVTTAFLYKCK 259  
 Qy 265 V-RGTAFLVFGQD--GEQISLPESLKRIPEDSGSEVLSRKVLGQVQNPRAEDLV 320  
 Db 260 VTADVITFTGREDLKQDKQEMQAMQNTMLINGIAQVTFDSETAKEISYVLEDLN 319  
 Qy 321 GKSLYSVATVILHSGSDMVAERSGIPVTSYQIHFTKTPKYPKPGMPFOLMVFVTPD 380  
 Db 320 NKLYIAVTVIESTGGFSEAEIPGIKYVLSYKLNLVATPLFLKPGIPYPIKQVKDSL 379

QY 381 GSPAYRVPVAVOGE-----DTQSLQO-GDGVAKLISINTHPSQKPLSITVTRTK 428  
Db 380 DQLVGGVPVFLNAQTDVNNQTSDLDPKSVTRVDDGVASFVNLPSGVTVLEPNVKTDA 439  
QY 429 QELSEAOATRWQALPYSTGVNSNNYLSLVLRTLPGETLVNFFLRMDRAHEAKIR 498  
Db 440 PDLPEENQAREGYRAIAYSLSQSYLYIDWTDNHHKALLVGEHLNI--IVTPKSPYIDKIT 497  
QY 489 YYYTLWNGRLILKAGQVRBPQDLVPLPSITTFIPSPFLVAYTYTLIGASQREWVA 548  
Db 498 HYNVLLSGKIIHFGTREKFSASVQSINIPVTQNMVPSKLLVLYIVITGEQ-TAELVS 556  
QY 549 DSVWVDKSCVGLVVKVSGSEDRQPVGQOMTLKIEGDHGARVLVAVDVGKGVFLNKK 608  
Db 557 DSVWLVIEKCGNQLQVHLSPADAYS-PGQTVSLNMATGMDSWVLAADVAVYGVQRG 615  
QY 609 NKLTSQKIWDVVEKADIGCTPGSGKDYAGVFSFAGLTFTSSGGQQTQRAELQCPQPAAR 668  
Db 616 AKKPLSERVFCLEKSLGCGAGGLNANVPHLAGLTFTLNANADDSQENDEPCKE-ILR 674  
QY 669 RRSVQLTEKMDKV-GKYPKE-LAKCCEDGNRENPMRFSQORTRTSLGEACKKVFELD 726  
Db 675 PRRTLO---KKIEEIAAKYKHSVVRKCCYDGCNVNDE-TCEQAAKRLSGPRCIFAPE 730  
QY 727 CCNYITELRRQARASHGLARSNLD-EDIIAEEINIVSRBFPSSWLNWVBDLKEPPKNG 785  
Db 731 CCVVASQLR---ANISHKDWQLGRLLHMKTLPLPVSKPEIRSYFPPSSWLMVEHLV---PRR- 783  
QY 786 ISYKLMNIEKDSITTEILAVMSDKGICVADPEVTVMQDFIDLRPLPSYVRNEOV 845  
Db 784 ---KQLOFALPDLSTLWEIQQIGISN-TGICVADTVKAKVFKVPLENNIPYSVVRGEQI 839  
QY 846 EIRAVLYNRQONQELKVRVELLHNPFCSLATTKRRHQQTIT-----IPPKSSLSVPY 898  
Db 840 QLKGTVNYR-TSGMVFVCKMGAVEGICTSESPVIDHQTKSKCVKRGKVGSSSHLVTF 898  
QY 899 VIPLKGTQGEVEKAAVHHIISDGRKSLVWEGIRMKNTVAVRTLDERLGRGVQ 958  
Db 899 TVLPLEIGHNINFSLEW--FGKEILVKTLLVPEGVKRESYSGV-TLDRPRGYIGISR 955  
QY 959 KEDIP---PADLSDDQVPTSETRILLQGTTPVAQMTEDAVDAERLKHILVTPSCGQNM 1015  
Db 956 KXEPFVRIPDL---VPKTEIKRILSVKGLLVGEILSAVLQEGINILTLHPKGSABEL 1012  
QY 1016 IGMTPTVIAVHLYDETEQNEKE---GLEKROGALIELIKKGYTQQLAFEPQPSAFAFVKR 1072  
Db 1013 MSVVPVYVYFVHLETGNHNI FHSDFPLEKQKLKXKLEGLMSIMSYNNADYSYVWKG 1072  
QY 1073 APSTWLTAYVVKVFLSALVNLIAIDSQLCGAVKMLLEKQPDGVFQEDAPVHQMIGG 1132  
Db 1073 SASTWLTAFALVGLQVKNVYEQNONSICNSLLWLVENYQLDNGSFKENSQYQIKLQGT 1132  
QY 1133 LR-NNNEKQMALTAFLVLSIQAKDICEQVNSLPGSITKAGDFLEANYNMLORSYVAI 1191  
Db 1133 LPVEARENSYLPFAFVIGIRKAFDIC--PLVKIDTALIKADNFLENTLPAQSTFTLAI 1190  
QY 1192 AGYALQMGRLKGLPLANKFLTAKOK-----NRWED---PGKQLYN-----VE 1231  
Db 1191 SAVALS-LGDKTHPQFERSVSKALKREALVKNPPIYRFWKONLQHKDSSVENTGTARWE 1249  
QY 1232 ATSVALLALQLKDFDPVVRVWLNQORYYGGVSGTQATFMVFOALAQOKDAPDHOE 1291  
Db 1250 TTYALLTSLNLDKINYNVPIVKNLSEORYGGGYSTQDTINAIEGLTELSLV---KK 1306  
QY 1292 LNLDSVLQPSRSSKITHRIHWEASLL-RSEETKENEGFTV-TAEGKQGTSLVMTYH 1349  
Db 1307 LRLSMIDVSYKHKGALHYKMKTDKNFLGRPEVLLNDDLLVSTGFGSLGTLVHTVTVH 1366  
QY 1350 AKAKDQLTGNKPLKTYIKPAETEKRPQDANKNTWILEICTRYRQDQATWS-----ILD 1404  
Db 1367 KTSISEVCS-FVLKTDITQDIEASHVRYGNSDYKRIIVACASYKPSRESSSGSHAVMD 1425

## RESULT 15

CSMS  
Complement C5 precursor - mouse  
N:Contains: C5a anaphylatoxin; C5b  
C:Species: Mus musculus (house mouse)  
C:Date: 19-Nov-1988 #sequence revision 15-Oct-1994 #text\_change 09-Jul-2004  
C:Accession: A35530; A27538; A40429  
R:Wetzel, R.A.; Fleischer, D.T.; Haviland, D.L.  
J. Biol. Chem. 265, 2435-2440, 1990  
A:Title: Deficiency of the murine fifth complement component (C5). A 2-base pair gene de  
A:Reference number: A35530; MUID:90153853; PMID:2303408  
A:Accession: A35530  
A:Molecule type: mRNA  
A:Residues: 1-215, 'L', <WET>  
A:Cross-references: UNIPROT:P06684; GB:M35526; GB:M35526; NID:g192302; PIDN:AAA37348.1;  
R:Wetzel, R.A.; Ogata, R.T.; Tack, B.F.  
Biochemistry 26, 737-743, 1987  
A:Title: Primary structure of the fifth component of murine complement.  
A:Reference number: A27538; MUID:87185363; PMID:2436653  
A:Accession: A27538  
A:Molecule type: mRNA  
A:Residues: 'PGI', 44-1680 <WET2>  
R:Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Wetzel, R.A.  
J. Biol. Chem. 266, 11818-11825, 1991  
A:Title: Structure of the murine fifth complement component (C5) gene. A large, highly i  
n component genes.  
A:Reference number: A40429; MUID:91268053; PMID:1711041  
A:Accession: A40429  
A:Molecule type: DNA  
A:Residues: 1-15 <HAV>  
A:Cross-references: GB:M64852  
C:Comment: Complement C5 contains two disulfide-linked chains, formed by removal of four  
(beta and alpha) chains.  
C:Comment: Activation of C5 initiates the spontaneous assembly of the late complement co  
is the foundation upon which the membrane attack complex is assembled.  
C:Comment: C5a has potent spasmogenic and chemotactic activity.  
C:Genetics:  
A:Map position: 2  
A:Introns: 22/3; 86/3; 140/3; 164/3; 195/2; 223/1; 253/2; 291/3; 334/1; 372/3; 434/3; 51  
3; 1224/1; 1292/3; 1343/3; 1364/3; 1392/1; 1411/2; 1445/3; 1470/3; 1506/1; 1534/1; 1564,  
C:Superfamily: alpha-2-macroglobulin  
C:Keywords: complement alternate pathway; complement pathway; cytolysis; glycoprotein;  
F:1-15/Domain: signal sequence #status predicted <SIG>  
F:19-674, 679-1679/Product: complement C5 #status predicted <CSB>  
F:19-674, 756-1679/Product: C5b #status predicted <CSB>  
F:19-674/Product: complement C5 and C5b beta chain #status predicted <CSBA>  
F:679-1679/Product: complement C5 alpha chain #status predicted <CSA>  
F:679-755/Product: C5a anaphylatoxin #status predicted <C5T>  
F:756-1679/Product: C5b alpha' chain #status predicted <C5BA>  
F:567-814, 635-670, 702-728, 703-735, 715-736, 870-1531, 1105-1163, 1379-1509, 1409-1478, 1524-1  
F:915, 1119, 1633/Binding site: carbohydrate (Asn) (covalent) #status predicted



**This Page Blank (uspto)**





Db 421 SITVTRTKQELSEACATRTMQUALPYSTVGNSSNNVHLHLSVLRTELRTGETLNVNPLRMD 480  
Qy 481 RAHEAKIRYTYLIMNKGRLKAGQVREPGQDLVPLSLTTFIPFISFRVAVYTLIGA 540  
Db 481 RAHEAKIRYTYLIMNKGRLKAGQVREPGQDLVPLSLTTFIPFISFRVAVYTLIGA 540  
Qy 541 SGQREVAVSVVDVVDKSCVGLSVKSGQSDQRPVPGQOMTLKIEGDHGVAVLVAVDK 600  
Db 541 SGQREVAVSVVDVVDKSCVGLSVKSGQSDQRPVPGQOMTLKIEGDHGVAVLVAVDK 600  
Qy 601 GVPVLMKKNLTKOSKIWDVWEKADIGCTPGSGKDVAGVPSDAGITFTSSSCQQAQRAEL 660  
Db 601 GVPVLMKKNLTKOSKIWDVWEKADIGCTPGSGKDVAGVPSDAGITFTSSSCQQAQRAEL 660  
Qy 661 QCQPQARRRRSQTTEKMDKVGKPKELKCCEDGKREMPNPFSCORRTFTSLGEAC 720  
Db 661 QCQPQARRRRSQTTEKMDKVGKPKELKCCEDGKREMPNPFSCORRTFTSLGEAC 720  
Qy 721 KKVFLDCCNVITELRQHARASHGLGARSNLDEDIIAENIVSRSEPPESHLWVDELKE 780  
Db 721 KKVFLDCCNVITELRQHARASHGLGARSNLDEDIIAENIVSRSEPPESHLWVDELKE 780  
Qy 781 PKXNGISTKLWNFLKDSITTEWILLAVNSDKKGCIVADPFEVTVMQDFFIDLRLPSVY 840  
Db 781 PKXNGISTKLWNFLKDSITTEWILLAVNSDKKGCIVADPFEVTVMQDFFIDLRLPSVY 840  
Qy 841 RNEQVEIRAVLYNQELKYRVVLLHNPAFCSLATTKRHQOQITIPPKSSLSVPIYI 900  
Db 841 RNEQVEIRAVLYNQELKYRVVLLHNPAFCSLATTKRHQOQITIPPKSSLSVPIYI 900  
Qy 901 VPLKTLGQEVKAAVYHFIISDGVKSLKVPEGRMKNVAVRTLDPRLRGEGVQKE 960  
Db 901 VPLKTLGQEVKAAVYHFIISDGVKSLKVPEGRMKNVAVRTLDPRLRGEGVQKE 960  
Qy 961 DIPADLSQVPTSETSEITLIGTPVQAOMTDAVDAERLKLIVTPSGCGEQNMIGWP 1020  
Db 961 DIPADLSQVPTSETSEITLIGTPVQAOMTDAVDAERLKLIVTPSGCGEQNMIGWP 1020  
Qy 1021 TVIAHYLDTEQWERFGLKQKQALELTKGYTQQLAFROPSSAFVRAFPSTWLTA 1080  
Db 1021 TVIAHYLDTEQWERFGLKQKQALELTKGYTQQLAFROPSSAFVRAFPSTWLTA 1080  
Qy 1081 YVVKVPSLAVNLTAIDSOVLGAVKWLILEKQKPGVFOEDAPVYHQEMICGLNNNEKD 1140  
Db 1081 YVVKVPSLAVNLTAIDSOVLGAVKWLILEKQKPGVFOEDAPVYHQEMICGLNNNEKD 1140  
Qy 1141 MALTAFLVLSIQBAKDCIEQVNSLPGSITKAGDFLEANYMNLQRSYTVATAGYALAQMG 1200  
Db 1141 MALTAFLVLSIQBAKDCIEQVNSLPGSITKAGDFLEANYMNLQRSYTVATAGYALAQMG 1200  
Qy 1201 RLKGPLLNKFLTTAKDKNWEDEPGKOLYNVEATSVALLALQLKDFDVPVFWLNEQR 1260  
Db 1201 RLKGPLLNKFLTTAKDKNWEDEPGKOLYNVEATSVALLALQLKDFDVPVFWLNEQR 1260  
Qy 1261 YGGGYSSTQATFWVQALAAQKQADPDHCELNLDVSLQLPSRSSKITHRIHWESASILLR 1320  
Db 1261 YGGGYSSTQATFWVQALAAQKQADPDHCELNLDVSLQLPSRSSKITHRIHWESASILLR 1320  
Qy 1321 SEETKENEGTVAAGKGGTISVWTHYAKAKDQITCNKEDLKVTIKPAPETEKRPQDA 1380  
Db 1321 SEETKENEGTVAAGKGGTISVWTHYAKAKDQITCNKEDLKVTIKPAPETEKRPQDA 1380  
Qy 1381 KNTWILEICRYRGDQDATMSILDISMGTGAPDITDQLKQANGVDVYISKYELDKAFSD 1440  
Db 1381 KNTWILEICRYRGDQDATMSILDISMGTGAPDITDQLKQANGVDVYISKYELDKAFSD 1440  
Qy 1441 RNTLIIYLDKVSHEDDCLAFKHQYFNVELTOPGAVKVAYNLEESTCFYFPEKEDG 1500  
Db 1441 RNTLIIYLDKVSHEDDCLAFKHQYFNVELTOPGAVKVAYNLEESTCFYFPEKEDG 1500  
Qy 1501 KLNKLCRDELRCRAEENCFIQKSDDDKVTLEERLDKACEFGVDVYVYKTRLVQVLSNDFDE 1560  
Db 1501 KLNKLCRDELRCRAEENCFIQKSDDDKVTLEERLDKACEFGVDVYVYKTRLVQVLSNDFDE 1560

Qy 1561 YVMAIEQTIKSGDEVQVGOORTFISPIKREALKLEEKHYLMGLSSDFWGEKPNLSY 1620  
Db 1561 YVMAIEQTIKSGDEVQVGOORTFISPIKREALKLEEKHYLMGLSSDFWGEKPNLSY 1620  
Qy 1621 IIGKDTVWEHWPFEDECOBENKQKQCDLGAFTESVMVFGCPN 1663  
Db 1621 IIGKDTVWEHWPFEDECOBENKQKQCDLGAFTESVMVFGCPN 1663

RESULT 2  
COS\_HUMAN  
ID COS\_HUMAN STANDARD; PRT; 1663 AA.  
AC P01024;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DE 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Complement C3 precursor [Contains: C3a anaphylatoxin].  
GN Name=C3;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85140166; PubMed=2579379;  
RA de Bruijn M.H.L., Fey G.H.;  
RT "Human complement component C3: cDNA coding sequence and derived  
RT primary structure.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:708-712 (1985).  
RN [2]  
RP SEQUENCE FROM N.A., AND VARIANTS GLY-102; PRO-314; LYS-863; ASP-1224  
RP AND THR-1367.  
RA Rieder M.J., Daniels R.L., da Ponte S.H., Hastings N.C., Ahearn M.O.,  
RA Rajkumar N., Yi Q., Nickerson D.A.;  
RT "SeattlesNPs, NHBI HLB682 program for genomic applications, UW-  
RT FHCR, Seattle, WA (URL: http://pga.gs.washington.edu).";  
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 672-748.  
RX MEDLINE=76069169; PubMed=1238393;  
RA Hugli T.E.;  
RT "Human anaphylatoxin (C3a) from the third component of complement.  
RT Primary structure.";  
RL J. Biol. Chem. 250:8293-8301 (1975).  
RN [4]  
RP SEQUENCE OF 955-966, AND SUBUNITS.  
RC TISSUE=Serum;  
RX MEDLINE=95293954; PubMed=7537971;  
RA Oxvig C., Haaning J., Kristensen L., Wagner J.M., Rubin I.,  
RA Stigbrand T., Gleich G.J., Sottrup-Jensen L.;  
RT "Identification of angiotensinogen and complement C3d as novel  
RT proteins binding the proform of eosinophil major basic protein in  
RT human pregnancy serum and plasma.";  
RL J. Biol. Chem. 270:13645-13651 (1995).  
RN [5]  
RP SEQUENCE OF 988-1036.  
RX MEDLINE=82174534; PubMed=6175959;  
RA Thomas M.L., Janatova J., Gray W.R., Tack B.F.;  
RT "Third component of human complement: localization of the internal  
RT thiolester bond.";  
RL Proc. Natl. Acad. Sci. U.S.A. 79:1054-1058 (1982).  
RN [6]  
RP SEQUENCE OF 1409-1563.  
RX MEDLINE=88154452; PubMed=3279119;  
RA Daoudaki M.E., Becherer J.D., Lambiris J.D.;  
RT "A 34-amino acid peptide of the third component of complement mediates  
RT properdin binding.";  
RL J. Immunol. 140:1577-1580 (1988).  
RN [7]  
RP STRUCTURE BY NMR OF C3A.  
RX MEDLINE=88276894; PubMed=3260670;  
RA Nettesheim D.G., Edalji R.P., Mollison K.W., Greer J.,

RA Zuideweg B.R.P.;  
 RT "Secondary structure of complement component C3a anaphylatoxin in  
 RT solution as determined by NMR spectroscopy: differences between  
 RT crystal and solution conformations.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:5036-5040 (1988).  
 [8]  
 RN MUTAGENESIS OF THIOESTER BOND REGION.  
 RX MEDLINE=92250565; PubMed=157777;  
 RA Isaac L., Isenman D.E.;  
 RT "Structural requirements for thioester bond formation in human  
 RT complement component C3. Reassessment of the role of thioester bond  
 RT integrity on the conformation of C3.";  
 RL J. Biol. Chem. 267:10062-10069 (1992).  
 [9]  
 RN DISULFIDE BONDS.  
 RX MEDLINE=93106233; PubMed=8416818;  
 RA Dolmer K., Sottrup-Jensen L.;  
 RT "Disulfide bridges in human complement component C3b.";  
 RL FEBS Lett. 315:85-90 (1993).  
 [10]  
 RN CARBOHYDRATE-LINKAGE SITE ASN-85.  
 RX MEDLINE=22660472; PubMed=12754519; DOI=10.1038/nbt827;  
 RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;  
 RT "Identification and quantification of N-linked glycoproteins using  
 RT hydrazide chemistry: stable isotope labeling and mass spectrometry.";  
 RL Nat. Biotechnol. 21:660-666 (2003).  
 [11]  
 RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 996-1303.  
 RX MEDLINE=98259089; PubMed=9596584;  
 RA Nagar B., Jones R.G., Diefenbach R.J., Isenman D.E., Rini J.M.;  
 RT "X-ray crystal structure of C3d: a C3 fragment and ligand for  
 RT complement receptor 2.";  
 RL Science 280:1277-1281 (1998).  
 [12]  
 RN VARIANT C3F/S.  
 RX MEDLINE=89309808; PubMed=2473125;  
 RA Ponansky M.C., Clissold P.M., Lachmann P.J.;  
 RT "The difference between human C3F and C3S results from a single amino  
 RT acid change from an asparagine to an aspartate residue at position  
 RT 1216 on the alpha-chain of the complement component, C3.";  
 RL J. Immunol. 143:1254-1258 (1989).  
 [13]  
 RN ERRATUM (RETRACTION).  
 RX MEDLINE=90063087; PubMed=2584723;  
 RA Ponansky M.C., Clissold P.M., Lachmann P.J.;  
 RL J. Immunol. 143:3860-3862 (1989).  
 [14]  
 RN VARIANTS GLY-102 AND PRO-314.  
 RX MEDLINE=91011240; PubMed=1976733;  
 RA Bosto M., Yong Pong K., So A.K., Koch C., Walport M.J.;  
 RT "Molecular basis of polymorphisms of human complement component C3.";  
 RL J. Exp. Med. 172:1011-1017 (1990).  
 [15]  
 RN VARIANT ASN-549.  
 RX MEDLINE=95050640; PubMed=7961791;  
 RA Singer L., Whitehead W.T., Akama H., Katz Y., Fishelson Z.,  
 RA Wiesel R.A.;  
 RT "Inherited human complement C3 deficiency. An amino acid substitution  
 RT in the beta-chain (Asp349 to Asn) impairs C3 secretion.";  
 RL J. Biol. Chem. 269:28494-28499 (1994).  
 [16]  
 RN VARIANT GLN-1320.  
 RA Watanabe Y., Matsui N., Yan K., Nishimukai H., Tokunaga K., Juji T.,  
 RA Kobayashi N., Kohsaka T.;  
 RT "A novel C3 allotype C3'F02' has an amino acid substitution that may  
 RT inhibit iC3b synthesis and cause C3-hypocomplementemia.";  
 RL Mol. Immunol. 30:62-62 (1993).  
 CC -!- FUNCTION: C3 plays a central role in the activation of the  
 CC complement system. Its processing by C3 convertase is the central  
 CC reaction in both classical and alternative complement pathways.  
 CC After activation C3b can bind covalently, via its reactive  
 CC thioester, to cell surface carbohydrates or immune aggregates.  
 CC -!- FUNCTION: Derived from proteolytic degradation of complement C3,

CC C3a anaphylatoxin is a mediator of local inflammatory process. It  
 CC induces the contraction of smooth muscle, increases vascular  
 CC permeability and causes histamine release from mast cells and  
 CC basophilic leukocytes.  
 CC -!- SUBUNIT: C3 precursor is first processed by the removal of 4 Arg  
 CC residues, forming two chains, beta and alpha, linked by a  
 CC disulfide bond. C3 convertase activates C3 by cleaving the alpha  
 CC chain, releasing C3a anaphylatoxin and generating C3b (beta chain  
 CC + alpha chain). During pregnancy, C3dg exists as a complex  
 CC (probably a 2:2:2 heterohexamer) with AGR and the proform of PRG2.  
 CC -!- PTM: C3b is rapidly split in two positions by factor I and a  
 CC cofactor to form iC3b (inactivated C3b) and C3f which is released.  
 CC Then iC3b is slowly cleaved (possibly by factor I) to form C3c and  
 CC C3dg. C3dg proteases produce other fragments such as C3d or C3g.  
 CC -!- POLYMORPHISM: There are two alleles: C3S (C3 slow), the most  
 CC common allele in all races and C3F (C3 fast), relatively frequent  
 CC in Caucasoids, less common in Black Americans, extremely rare in  
 CC Orientals.  
 CC -!- DISEASE: Defects in C3 are the cause of C3 deficiency  
 CC (MIM:120700). It can result in susceptibility to pyogenic  
 CC infection.  
 CC -!- SIMILARITY: Contains 1 anaphylatoxin-like domain.  
 CC -!- SIMILARITY: Contains 1 NTR domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; K02765; AAA85332.1; ..  
 DR EMBL; AY513239; AAR89906.1; ..  
 DR PIR; A94065; C3HU.  
 DR PDB; 1C3D; X-ray; @=-.  
 DR PDB; 1GHQ; X-ray; A=994-1300.  
 DR SWISS-2DPAGE; P01024; HUMAN.  
 DR Sienna-2DPAGE; P01024; ..  
 DR Genew; HGNC:1318; C3.  
 DR MIM; 120700; ..  
 DR GO; GO:0005102; F:receptor binding; TAS.  
 DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; TAS.  
 DR GO; GO:0006955; P:immune response; TAS.  
 DR GO; GO:0007165; P:signal transduction; TAS.  
 DR InterPro; IPR002890; A2M\_N.  
 DR InterPro; IPR009048; AM\_Receptor\_bind.  
 DR InterPro; IPR000020; Anaphylatoxin.  
 DR InterPro; IPR001840; Anaphylatoxin.  
 DR InterPro; IPR008964; Invasin\_intimin.  
 DR InterPro; IPR001599; MacroglobulinA2.  
 DR InterPro; IPR001134; Netrin\_C.  
 DR InterPro; IPR008930; Terp\_cyc\_toroid.  
 DR InterPro; IPR008993; TIME\_like.  
 DR Pfam; PF00207; A2M\_1.  
 DR Pfam; PF01835; A2M\_N; 1.  
 DR Pfam; PF01821; ANATO; 1.  
 DR Pfam; PF01759; NTR; 1.  
 DR PRINTS; PR00004; ANAPHYLATOXN.  
 DR ProDom; PD003264; Anaphylatoxin; 1.  
 DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.  
 DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.  
 DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.  
 DR PROSITE; PS00189; NTR; 1.  
 KW 3D-structure; Complement alternate pathway; Complement pathway;  
 KW Direct protein sequencing; Disease mutation; Glycoprotein;  
 Query Match 99.9%; Score 8598; DB 1; Length 1663;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1661; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MGSTGSPSLLLLLTHPLALGSPKYSITNILESEFTWLEHDAQGVVTVVH 60  
 |||||

Db 1 MGTSGPSLLLLLTHLPALGSPMYSIITPNILBLESEETWLEAHDAQGVVTVVH 60  
Qy 61 DFGKKLVLSSEKTLVTPATNMGVNTFTIPANREFKSEKGRNFVTVQATFGTVVVKY 120  
Db 61 DFGKKLVLSSEKTLVTPATNMGVNTFTIPANREFKSEKGRNFVTVQATFGTVVVKY 120  
Qy 121 VLVSQSGYLFOTDXTIYTPGASTVLYRIFTVNHKLLPVGTVMVNIENPEGIPVKQDSL 180  
Db 121 VLVSQSGYLFOTDXTIYTPGASTVLYRIFTVNHKLLPVGTVMVNIENPEGIPVKQDSL 180  
Qy 181 SSQNLGVLPLSDIPELVNMGMQKIRAYENSPQVFESEFVEKRYVLPSEFIVEPTE 240  
Db 181 SSQNLGVLPLSDIPELVNMGMQKIRAYENSPQVFESEFVEKRYVLPSEFIVEPTE 240  
Qy 241 KFYIYNEKGLVTTITARFLYKKGVEGTAFLVFGTQDGEQRIISLPESIKRPIEDGSGEV 300  
Db 241 KFYIYNEKGLVTTITARFLYKKGVEGTAFLVFGTQDGEQRIISLPESIKRPIEDGSGEV 300  
Qy 301 VLSRKVLGQVQNPRAEDLVGSLVYSATVILHSGSDMVQAERSGIPVTSPIYHPTKT 360  
Db 301 VLSRKVLGQVQNPRAEDLVGSLVYSATVILHSGSDMVQAERSGIPVTSPIYHPTKT 360  
Qy 361 PKYFKGMPFDLMVFTNPDGSPAYRVPVAVOGEDTVQSLTQDGVAKLSINTHPSOKPL 420  
Db 361 PKYFKGMPFDLMVFTNPDGSPAYRVPVAVOGEDTVQSLTQDGVAKLSINTHPSOKPL 420  
Qy 421 SITVRTKQELSEAEQATRTMOALPYSTVGNSSNNYLHLSVLRTELPGETLNVNFLRMD 480  
Db 421 SITVRTKQELSEAEQATRTMOALPYSTVGNSSNNYLHLSVLRTELPGETLNVNFLRMD 480  
Qy 481 RAHEAKIRYTYLIMNKGRLKAGROVREPQDGLVPLSITTFIPFSRLVAVYTLIGA 540  
Db 481 RAHEAKIRYTYLIMNKGRLKAGROVREPQDGLVPLSITTFIPFSRLVAVYTLIGA 540  
Qy 541 SGQREVYADSVVDVKDSVGVSLVKSQSDRQVPVQGMQLKIEGDHGARVVLVAVDK 600  
Db 541 SGQREVYADSVVDVKDSVGVSLVKSQSDRQVPVQGMQLKIEGDHGARVVLVAVDK 600  
Qy 601 GVFLVANKKLTOSKTDVVEKADIGCTPGSGKDVAGVPSDAGLTFTSSSQOQTAQRAEL 660  
Db 601 GVFLVANKKLTOSKTDVVEKADIGCTPGSGKDVAGVPSDAGLTFTSSSQOQTAQRAEL 660  
Qy 661 QCPQPAARRRSVOLTKEKMDKVGYPKELRKCCEGDMRENPMRPFSCQRTTRFISLGEAC 720  
Db 661 QCPQPAARRRSVOLTKEKMDKVGYPKELRKCCEGDMRENPMRPFSCQRTTRFISLGEAC 720  
Qy 721 KKVFLDCCNVITELRRQHARASHGLGARSNLDEIIAENIVSRSEPPESLWNVDELXE 780  
Db 721 KKVFLDCCNVITELRRQHARASHGLGARSNLDEIIAENIVSRSEPPESLWNVDELXE 780  
Qy 781 PPKNGISTKLWNTFLKDSITTWELIAVMSDKKGCIVADPFEVTVMQDFFIDLRPLPSV 840  
Db 781 PPKNGISTKLWNTFLKDSITTWELIAVMSDKKGCIVADPFEVTVMQDFFIDLRPLPSV 840  
Qy 841 RNQVEIRAVLYNRQNELKVRVELLHNPACSLATTKRHOQITTPPKSSLSVPYVI 900  
Db 841 RNQVEIRAVLYNRQNELKVRVELLHNPACSLATTKRHOQITTPPKSSLSVPYVI 900  
Qy 901 VPLKTLQVEVEVAAVHHFISDGVKSLKVVPEGIRMNKTVAVRTLDPERLREGVQKE 960  
Db 901 VPLKTLQVEVEVAAVHHFISDGVKSLKVVPEGIRMNKTVAVRTLDPERLREGVQKE 960  
Qy 961 DIPADLSDOVPTSETRILLOGTTPVQMTEDAVDAERLKHILVTPSGCGEQNWGMT 1020  
Db 961 DIPADLSDOVPTSETRILLOGTTPVQMTEDAVDAERLKHILVTPSGCGEQNWGMT 1020  
Qy 1021 TVTAVHVLDETEQWFKGLEKROGAELELIKGYTQQLAFRQPSAFAAFVGRAPSTWLTA 1080  
Db 1021 TVTAVHVLDETEQWFKGLEKROGAELELIKGYTQQLAFRQPSAFAAFVGRAPSTWLTA 1080  
Qy 1081 YVVKVFLAVNLTAIDSQVLCGAVKWLILEKQKPGVFOEDAPVTHQEMIGGLNNNEKD 1140  
Db 1081 YVVKVFLAVNLTAIDSQVLCGAVKWLILEKQKPGVFOEDAPVTHQEMIGGLNNNEKD 1140

## RESULT 3

CO3\_RAT STANDARD; PRT; 1663 AA.  
ID CO3\_RAT STANDARD; PRT; 1663 AA.  
AC F01026;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Complement C3 precursor [Contains: C3a anaphylatoxin].  
GN Name=C3;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OC NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Wistar; TISSUE=Liver;  
RX MEDLINE=90245672; PubMed=2336397;  
RA Misumi Y., Shoda M., Ikehara Y.;  
RT "Nucleotide and deduced amino acid sequence of rat complement C3.";  
RL Nucleic Acids Res. 18:2178-2178(1990).  
RN [2]  
RP SEQUENCE OF 671-748.  
RX MEDLINE=75062262; PubMed=309768;  
RA Jacobs J.W., Rubin J.S., Hugli T.E., Bogardt R.A., Mariz I.K.,  
RA Daniels J.S., Daughaday W.H., Bradshaw R.A.;  
RT "Purification, characterization, and amino acid sequence of rat anaphylatoxin (C3a).";  
RL Biochemistry 17:5031-5038 (1978).  
RN [3]  
RP SEQUENCE OF 1316-1595 FROM N.A.  
RX MEDLINE=89380332; PubMed=2674144;  
RA Sundstrom S.A., Komm B.S., Ponce-De-Leon H., Yi Z., Teuscher C.,  
RA Lyttle C.R.;  
RT "Estrogen regulation of tissue-specific expression of complement C3.";  
RL J. Biol. Chem. 264:16941-16947(1989).

CC -!- FUNCTION: C3 plays a central role in the activation of the complement system. Its processing by C3 convertase is the central reaction in both classical and alternative complement pathways. After activation C3b can bind covalently, via its reactive thioester, to cell surface carbohydrates or immune aggregates.

CC -!- FUNCTION: Derived from proteolytic degradation of complement C3, C3a anaphylatoxin is a mediator of local inflammatory process. It induces the contraction of smooth muscle, increases vascular permeability and causes histamine release from mast cells and basophilic leukocytes.

CC -!- SUBUNIT: C3 precursor is first processed by the removal of 4 Arg residues, forming two chains, beta and alpha, linked by a disulfide bond. C3 convertase activates C3 by cleaving the alpha chain, releasing C3a anaphylatoxin and generating C3b (beta chain + alpha' chain).

CC -!- SIMILARITY: Contains 1 anaphylatoxin-like domain.

CC -!- SIMILARITY: Contains 1 NTR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X52477; CAA36716.1; -

DR EMBL; M29866; AAB40837.1; ALT\_SEQ.

DR PIR; S15764; C3RT.

DR PDB; 1QQF; X-ray; A=1010-1286.

DR PDB; 1QSG; X-ray; A/B/C/D=1010-1286.

DR RGD; 2232; C3.

DR InterPro; IPR002890; A2M\_N.

DR InterPro; IPR009048; AM\_Receiver\_bind.

DR InterPro; IPR000020; Anaphylatoxin.

DR InterPro; IPR001840; Anaphylatoxin.

DR InterPro; IPR008964; Invasion\_inhibitor.

DR InterPro; IPR001599; MacroglobulinA2.

DR InterPro; IPR001134; Netrin\_C.

DR InterPro; IPR008930; Terp\_cyc\_toroid.

DR InterPro; IPR008993; TIMP\_like.

DR Pfam; PF00207; A2M; 1.

DR Pfam; PF01835; A2M\_N; 1.

DR Pfam; PF01821; ANATO; 1.

DR Pfam; PF01755; NTR; 1.

DR PRINTS; PRO00004; ANAPHYLATOXIN.

DR PRODOM; PD03264; Anaphylatoxin; 1.

DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.

DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.

DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.

DR PROSITE; PS01189; NTR; 1.

KW 3D-structure. Complement alternate pathway; Complement pathway;

KW Direct protein sequencing; Glycoprotein; Inflammatory response;

KW Plasma; Signal; Thioester bond.

FT SIGNAL 1 24

FT CHAIN 25 1663 Complement C3.

FT CHAIN 25 666 Complement C3 beta chain.

FT CHAIN 671 1663 Complement C3 alpha chain.

FT PEPTIDE 671 748 C3a anaphylatoxin.

FT CHAIN 749 1663 Complement C3b alpha' chain.

FT DOMAIN 693 728 Anaphylatoxin-like.

FT DOMAIN 1518 1661 NTR.

FT SITE 748 749 Cleavage (by C3 convertase).

FT DISULFID 558 816 Interchain (by similarity).

FT DISULFID 626 661 By similarity.

FT DISULFID 693 720 By similarity.

FT DISULFID 694 727 By similarity.

FT DISULFID 707 728 By similarity.

FT DISULFID 873 1513 By similarity.

FT DISULFID 1101 1158 By similarity.

FT DISULFID 1358 1489 By similarity.

FT DISULFID 1389 1458 By similarity.

FT DISULFID 1506 1511 By similarity.

FT DISULFID 1518 1590 By similarity.

FT DISULFID 1537 1661 By similarity.

FT CROSSLINK 1010 1013 N-linked (GlcNAc...) (Probable).

FT CARBOHYD 939 939 N-linked (GlcNAc...) (Probable).

FT CARBOHYD 1617 1617 N-linked (GlcNAc...) (Probable).

FT CONFLICT 721 722 LX -> KL (in Ref. 2).

FT TURN 1011 1012

FT HELIX 1013 1031

FT TURN 1032 1032

FT TURN 1032 1032

FT HELIX 1034 1037

FT HELIX 1039 1041

FT HELIX 1042 1057

FT TURN 1058 1059

FT STRAND 1060 1060

FT TURN 1062 1063

FT STRAND 1066 1066

FT TURN 1070 1071

FT HELIX 1076 1089

FT TURN 1090 1092

FT HELIX 1097 1111

FT STRAND 1112 1112

FT TURN 1114 1115

FT STRAND 1118 1118

FT HELIX 1127 1134

FT TURN 1137 1138

FT HELIX 1139 1158

FT TURN 1159 1161

FT TURN 1163 1164

FT HELIX 1165 1180

FT TURN 1181 1182

FT HELIX 1186 1198

FT TURN 1199 1200

FT TURN 1204 1205

FT HELIX 1206 1213

FT STRAND 1215 1215

FT TURN 1216 1218

FT STRAND 1219 1219

FT TURN 1223 1224

FT HELIX 1226 1242

FT TURN 1243 1244

FT TURN 1246 1247

FT HELIX 1249 1258

FT TURN 1263 1264

FT TURN 1266 1267

FT HELIX 1269 1285

SQ SEQUENCE 1663 AA; 186460 MW; 2F87CCB143CDD4BC CRC64;

Query Match 79.3%; Score 6825; DB 1; Length 1663;

Best Local Similarity 78.1%; Pred. No. 0;

Matches 1301; Conservative 171; Mismatches 188; Indels 6; Gaps 5;

Qy 1 MGTSQPSL--LLLLTHLPALGSPMYSIITPNILRLSESETVLEAHDAGQDPVTVT 58

Db 1 MGFTSGQLLVLLLLSLALGSPMYSIITPNVLRLESEETFILEAHDAGQDPVTVT 60

Qy 59 VHDFFGKLVLSSEKTVLTPATNHMGVNTFTIPANREFKSEKRNKFTVTCATGCTQWE 118

Db 61 VQDFL-KKQVLTSEKTVLTGATGLNRVFIKIPASKEFNADKG-HKYVVVANTGATWE 118

Qy 119 KVVVLSQSGYLPIDTKTITTPGSTVLYRFTVNNHKLIPVGRVNMNIENPEGIPKQD 178

Db 119 KAVLVSFQSGYLPIDTKTITTPGSTVLYRFTVNNHKLIPVGRVNMNIENPEGIPKRD 178

Qy 179 SLSSQNLQGLVPLSWDIPELVNMGOWKIRAYVENSPOQVFTSEFEVKEVLPSEVIVEP 238

Db 179 ILSSHQVGLPLSNWIPELVNMGOWKIRAFVHAPKQTFSAEFVKEVLPSEVIVEP 238

Qy 239 TEKFYIYNEKGLVTTITARFLYKKGEGTAFVIFGQQGQRISLPESLKRIFIEGSG 298

Db 239 TEKFYIHGPKGLEVSITARFLYKKNVDGTAFAVFGVQDEKISLALSTRVIEDGSG 298

Qy 299 EVVLSKVLDDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQASRGIPVITSFYQHFT 358

Db 299 EAVLSRKVLMDGVRPSPEALVGKSLVSVTVILHSGSDMVEABSGIPVITSPVQIHFT 358  
 Qy 359 KTKPKYKPGVPFDMVFTNPDGSPAYRVPVAVQGEDTVQSLTQDGGVAKLSINTHPSQK 418  
 Db 359 KTKPKYKPGVPFDMVFTNPDGSPAYRVPVAVQGEDTVQSLTQDGGVAKLSINTHPSQK 417  
 Qy 419 PLSTVETKQELSEABQATRTMQALPYSTVGNSSNNYLHLSVRLTELRPGSTLNNFLR 478  
 Db 418 PLSTVETKQELSEABQATRTMQALPYSTVGNSSNNYLHLSVRLTELRPGSTLNNFLR 477  
 Qy 479 MDRAHEAKIRYTYLIMNKGRLLKAGQVREPGQDLVLPISITTDFTPSRLVAVYTLI 538  
 Db 478 TDAGEAKIRYTYLIMNKGRLLKAGQVREPGQDLVLPISITTDFTPSRLVAVYTLI 537  
 Qy 539 GASQOREVADSVWVDKSGVSLVVKSGQSEDRQVPVPGQMTLKIISGDHARVLVAV 598  
 Db 538 GASQOREVADSVWVDKSGVSLVVKSGQSEDRQVPVPGQMTLKIISGDHARVLVAV 597  
 Qy 599 DKGVFVLNKKNKLTQSKIWDVVEKADIGCTPGSGKDYAGVPSDAGLTFTSSSGQOTQARA 658  
 Db 598 DKGVFVLNKKNKLTQSKIWDVVEKADIGCTPGSGKDYAGVPSDAGLTFTSSSGQOTQARA 657  
 Qy 659 ELQCPQPAARRRSVOLTEKMDKGVKY-PKELRKCCEGDGKRENMRFSCORRTFRISLG 717  
 Db 658 DPECAKPAARRRSVOLTEKMDKGVKY-PKELRKCCEGDGKRENMRFSCORRTFRISLG 717  
 Qy 718 EACKVPLDCNVITELRCHRAASHGLARNDLEDIIAENIVSRSEFFESLMWVED 777  
 Db 718 EACKVPLDCNVITELRCHRAASHGLARNDLEDIIAENIVSRSEFFESLMWVED 777  
 Qy 778 LKEPKNGISTKLMNIFLKDSTITTEILAVSDSKGICVADPEFTVWQDFFIDRLPY 837  
 Db 778 LKEPKNGISTKLMNIFLKDSTITTEILAVSDSKGICVADPEFTVWQDFFIDRLPY 837  
 Qy 838 SVRNEQVEIRAVLYNRYONKELKRVVLLHNPAPCSLATTKRHQQTITTPPKSSLSVP 897  
 Db 838 SVRNEQVEIRAVLYNRYONKELKRVVLLHNPAPCSLATTKRHQQTITTPPKSSLSVP 897  
 Qy 898 YVIVLTKGLOEVEVAAYVHPIISGVKPSLKVYPEGIRNMTKTVAVTILDEPLRGEGV 957  
 Db 898 YVIVLTKGLOEVEVAAYVHPIISGVKPSLKVYPEGIRNMTKTVAVTILDEPLRGEGV 957  
 Qy 958 QKEDIPPADLSQVPTSETRILLQGTPEVAQMTDAVDAERLKHILVTPSGCGEQNWIG 1017  
 Db 958 QKEDIPPADLSQVPTSETRILLQGTPEVAQMTDAVDAERLKHILVTPSGCGEQNWIG 1017  
 Qy 1018 MTPPTVIAVHLDTEQWEKLEKQGALELTKKGTQQLAFROPSSAFAPVRAFTW 1077  
 Db 1018 MTPPTVIAVHLDTEQWEKLEKQGALELTKKGTQQLAFROPSSAFAPVRAFTW 1077  
 Qy 1078 LTAIVVVKFSLAVNLIAIDSLQCGAVKWLILKQKPGVQEDAPVHQEMIGGLRNN 1137  
 Db 1078 LTAIVVVKFSLAVNLIAIDSLQCGAVKWLILKQKPGVQEDAPVHQEMIGGLRNN 1137  
 Qy 1138 EADMALTAFLVLSIQEAKDICEOVNLSFGSTKAGDLEANYNLOSRYVATAGVALA 1197  
 Db 1138 EADMALTAFLVLSIQEAKDICEOVNLSFGSTKAGDLEANYNLOSRYVATAGVALA 1197  
 Qy 1198 QMGLKGPLANKLFTTAKDNKRWEDPGKOLYNVEATSVALLALQLKDFDVPVPPVRLN 1257  
 Db 1198 QMGLKGPLANKLFTTAKDNKRWEDPGKOLYNVEATSVALLALQLKDFDVPVPPVRLN 1257  
 Qy 1258 EORYGGYGSTQATFWFQALAOYOKDAPDHOENLVDVLSQLPSRSSKITHRWESAS 1317  
 Db 1258 EORYGGYGSTQATFWFQALAOYOKDAPDHOENLVDVLSQLPSRSSKITHRWESAS 1317  
 Qy 1318 LLRSEETKENGFTVTAEGKQGTLSVVTWTHAKAKOQTCNFKDLKVTIKPAPETKRP 1377  
 Db 1318 LLRSEETKENGFTVTAEGKQGTLSVVTWTHAKAKOQTCNFKDLKVTIKPAPETKRP 1377  
 Qy 1378 QDAKNTMILICTRYRGDODATMSILDISMGTGAPDITDLKLANGVDRIYSKYELDKA 1437  
 Db 1378 QDAKNTMILICTRYRGDODATMSILDISMGTGAPDITDLKLANGVDRIYSKYELDKA 1437

Qy 1438 FSDRNTLIIYLDKVSHEDDCLAFKVHOFYFVELIQPGAVKYVAYNLEESCTRYHPEK 1497  
 Db 1438 FSDRNTLIIYLDKVSHEDDCLAFKVHOFYFVELIQPGAVKYVAYNLEESCTRYHPEK 1497  
 Qy 1498 EDGKLNLCRDELCRAEENCFIQKSDDKVTLLEERLDKACBPGVDYVYKTLVKVQSLND 1557  
 Db 1498 EDGKLNLCRDELCRAEENCFIQKSDDKVTLLEERLDKACBPGVDYVYKTLVKVQSLND 1557  
 Qy 1558 FDEVIMAEOTIKSGSDEVQVGOORTFISPIKREALKLEEKHVMGLSGDSDFWGEKPN 1617  
 Db 1558 FDEVIMAEOTIKSGSDEVQVGOORTFISPIKREALKLEEKHVMGLSGDSDFWGEKPN 1617  
 Qy 1618 LSYIIGKDTWVEHWPBEDEQDSEENKQKQDILGAFTESWVFGCPN 1663  
 Db 1618 LSYIIGKDTWVEHWPBEDEQDSEENKQKQDILGAFTESWVFGCPN 1663

RESULT 4  
 CO3\_CAVPO  
 ID CO3\_CAVPO STANDARD; PRT; 1666 AA.  
 AC P12387;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Complement C3 precursor [Contains: C3a anaphylatoxin].  
 GN Name=C3;  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
 OX NCBI\_TaxID=10141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90307996; PubMed=1973176;  
 RA Auerbach H.S., Burger R., Dodds A., Colten H.R.;  
 RT "Molecular basis of complement C3 deficiency in guinea pigs.";  
 RL J. Clin. Invest. 86:96-106(1990).  
 RN [2]  
 RP SEQUENCE OF 576-753.  
 RX MEDLINE=89113342; PubMed=3064079;  
 RA Gerard N.P., Lively M.O., Gerard C.;  
 RT "Amino acid sequence of guinea pig C3a anaphylatoxin.";  
 RL Protein Seq. Data Anal. 1:473-478(1988).  
 RN [3]  
 RP SEQUENCE OF 993-1032.  
 RX MEDLINE=8317889; PubMed=6838833;  
 RA Thomas M.L., Tack B.F.;  
 RT "Identification and alignment of a thiol ester site in the third component of guinea pig complement.";  
 RL Biochemistry 22:942-947(1983).  
 CC -!- FUNCTION: C3 plays a central role in the activation of the complement system. Its processing by C3 convertase is the central reaction in both classical and alternative complement pathways. After activation C3b can bind covalently via its reactive thiolester to cell surface carbohydrates or immune aggregates. CC  
 CC -!- FUNCTION: Derived from proteolytic degradation of complement C3, C3a anaphylatoxin is a mediator of local inflammatory process. It induces the contraction of smooth muscle, increases vascular permeability and causes histamine release from mast cells and basophilic leukocytes. CC  
 CC -!- SUBUNIT: C3 precursor is first processed by the removal of 4 Arg residues, forming two chains, beta and alpha, linked by a disulfide bond. C3 convertase activates C3 by cleaving the alpha chain, releasing C3a anaphylatoxin and generating C3b (beta chain + alpha' chain). CC  
 CC -!- SIMILARITY: Contains 1 anaphylatoxin-like domain.  
 CC -!- SIMILARITY: Contains 1 NTR domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. CC  
 CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial





QY	1254	RHLNQRYYGGYGSGTQTATFWFQAALAQYKQADAPDHQELNLVSLQLPSRSKTHRIHW	1313
Dd	1259	RWLNEQRYYRGYGGSTQTATFWVFQAALAQYOTDVPFDHKLJNMVEVALQLPSRSPSKFLAVW	1318
QY	1314	ESASLLRSEETKENEGFTVTAEKGQGTLSVVMTMHAKAQQOLTCNKFDLVKITKPAPET	1373
Dd	1319	EAGSLLRSEATKQNEGFKLTAKGKGQGLSVAVVYAKTKRKVCCKNFDLAVTLKPAPDT	1378
QY	1374	EKRPODAKNWTMLIRICTRYBGDDATMSILDISMTGPAPDITDDLKQLANGVDVYSKYE	1433
Dd	1379	VKKQEAEXTMILIGICTRYLGDQDATMSILDISMTGFIPDITDDUKLATGVDRYISKYE	1438
QY	1434	LDKAFSDRNLTIIIVLDKVSHSEDCTAFKHQHFNVNELIQGAVKIAYYNLEBSCTRFF	1493
Dd	1439	MWKDFS-KNTLIIVLDKVSHSEBCLSFKTHOFFNVGLIQGSVKIYYNLDSTCTQFY	1497
QY	1494	HPEKEDGKLNKLCDELRCRAEBNCITQKSDDKVTLEERLDKACBPVGVDVYVKRLVKVQ	1553
Dd	1498	HPEKEDGMNLKCHKDLRCRAEENCFIQ-LPEKITLDERLEKACEPGVDVYVKTCLKWE	1556
QY	1554	LSNDFDEYIMAIETIYSGSDVQVQORTTFISPIKCEALKLBKHYLMWGSSDPFG	1613
Dd	1557	JSDDFDEYIMTIEQVIKSGDVEQAGKERFISHIXCRDALHXEGHYLMWGSSDLWG	1616
QY	1614	EKPNSLYIIGKDTVWEHPBEDECQDEENQKQCQIDLGAFTESMYVFGCPN	1663
Dd	1617	ERPNNSYIIGKDTVWEAWPAERBCQDEENQKQCQIDLTFTENMYVFGCPN	1666

RESULT 5

ID	Q80XP1	PRELIMINARY;	PRT; 1663 AA.
AC	Q80XP1		
DT	01-JUN-2003	(TREMBUrel. 24, Created)	
DT	01-JUN-2003	(TREMBUrel. 24, Last sequence update)	
DT	01-MAR-2004	(TREMBUrel. 26, Last annotation update)	
DE		Complement component 3.	
OS		Mus musculus (Mouse).	
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OX		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
NCBI		_TaxID=10090;	
RP	[1]	SEQUENCE FROM N.A.	
RC	STRAIN=FVB/N;	TISSUE=Liver;	
RX	MEDLINE=23386257;	PubMed=12477932;	
RA	Strausberg R.U., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M.J., Soares M.B., Bonaldi M.P., Casavant T.L., Scheetz T.E.,		
RA	Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,		
RA	Posk S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale T., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,		
RA	Fahy J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzywinski A.I., Skalska U., Smalley D.E., Schnerch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FVB/N;	TISSUE=Liver;	
RA	Strausberg R.;		
RL	Submitted (Jan.-2003) to the EMBL/GenBank/DBEP databases.		
DR	EMBL; BC043338; AAA43338.1; -		
DR	HSSP; P01026; 1QQF.		



CC residues, forming two chains, beta and alpha, linked by a  
CC disulfide bond. C3 convertase activates C3 by cleaving the alpha  
CC chain, releasing C3a anaphylatoxin and generating C3b (beta chain  
CC + alpha' chain).

CC -|- ALTERNATIVE PRODUCTS:  
CC Event=Alternative initiation;  
CC Comment=2 isoforms, Long (shown here) and Short, are produced by  
CC alternative initiation;  
CC -|- PMW: C3b is rapidly split in two positions by factor I and a  
CC cofactor to form iC3b (inactivated C3b) and C3f which is released.  
CC Then iC3b is slowly cleaved (possibly by factor I) to form C3c and  
CC C3g. Other proteases produce other fragments such as C3d or C3g.  
CC -|- SIMILARITY: Contains 1 anaphylatoxin-like domain.  
CC -|- SIMILARITY: Contains 1 NTR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL: K02782; AAC42013.1; -;  
CC EMBL: J00369; AAA37336.1; -;  
CC EMBL: J00367; AAA37336.1; JOINED.  
CC EMBL: M3032; AAA37378.1; -;  
CC EMBL: Z37998; CAA86099.2; -;  
CC PIR: A92459; C3MS.  
CC HSHP: P01026; 1QOF.  
CC MG: MGI:88227; C3.  
CC GO: GO:0006954; P:inflammatory response; IMP.  
CC GO: GO:0050766; P:positive regulation of phagocytosis; IMP.  
CC InterPro: IPR002890; A2M\_N.  
CC InterPro: IPR003048; A2M\_N.  
CC InterPro: IPR000020; Anaphylatoxin.  
CC InterPro: IPR001840; Anaphylatoxin.  
CC InterPro: IPR008964; Invasin intimin.  
CC InterPro: IPR001593; MacroglobulinA2.  
CC InterPro: IPR001134; Netrin\_C.  
CC InterPro: IPR008930; Terp cyc toroid.  
CC InterPro: IPR008993; TIMP\_like.  
CC Pfam: PF00207; A2M; 1.  
CC Pfam: PF01835; A2M\_N; 1.  
CC Pfam: PF01821; ANATO; 1.  
CC Pfam: PF01759; NTR; 1.  
CC PRINTS: PR00004; ANAPHYLATOXN.  
CC ProDom: PD003264; Anaphylatoxin; 1.  
CC PROSITE: PS00477; ALPHA\_2\_MACROGLOBULIN; 1.  
CC PROSITE: PS01177; ANAPHYLATOXIN\_1; 1.  
CC PROSITE: PS01178; ANAPHYLATOXIN\_2; 1.  
CC PROSITE: PS01019; NTR; 1.  
CC PROSITE: PS01019; NTR; 1.  
KW Alternative initiation; Complement alternate pathway;  
KW Complement pathway; Direct protein sequencing; Glycoprotein;  
KW Inflammatory response; Plasma; Signal; Thioester bond.

FT SIGNAL 1 24  
FT CHAIN 25 1663 Complement C3, isoform Long.  
FT CHAIN 1129 1663 Complement C3, isoform Short.  
FT INIT MET 1129 1129 For isoform Short.  
FT CHAIN 25 666 Complement C3 beta chain.  
FT CHAIN 671 1663 Complement C3 alpha chain.  
FT CHAIN 671 1663 C3a anaphylatoxin.  
FT PEPTIDE 749 1663 Complement C3b alpha' chain.  
FT CHAIN 749 954 Complement C3c fragment.  
FT CHAIN 955 1303 Complement C3dg fragment.  
FT CHAIN 955 1001 Complement C3g fragment.  
FT CHAIN 1002 1303 Complement C3d fragment.  
FT PEPTIDE 1304 1320 C3f fragment.  
FT DOMAIN 693 728 Anaphylatoxin-like.  
FT SITE 1518 1661 NTR.  
FT SITE 748 749 Cleavage (by C3 convertase).  
FT SITE 1303 1304 Cleavage (by factor I).  
FT SITE 1320 1321 Cleavage (by factor I).

FT DISULFID 559 816 Interchain (By similarity).  
FT DISULFID 626 661 By similarity.  
FT DISULFID 693 720 By similarity.  
FT DISULFID 694 727 By similarity.  
FT DISULFID 707 728 By similarity.  
FT DISULFID 873 1513 By similarity.  
FT DISULFID 1101 1158 By similarity.  
FT DISULFID 1358 1489 By similarity.  
FT DISULFID 1389 1458 By similarity.  
FT DISULFID 1506 1511 By similarity.  
FT DISULFID 1518 1590 By similarity.  
FT DISULFID 1537 1661 By similarity.  
FT DISULFID 1637 1646 By similarity.  
FT CARBOHYD 939 939 N-linked (GlcNAc...).  
FT CARBOHYD 1617 1617 N-linked (GlcNAc...).  
FT CROSSLINK 1010 1013 Iso-glutamyl cysteine thioester (Cys-Gln)  
FT SEQUENCE 1663 AA; 166482 MW; D55546CC769BEA19 CRC64;  
Query Match 78.4%; Score 6749; DB 1; Length 1663;  
Best Local Similarity 77.1%; Pred. No. 0;  
Matches 1284; Conservative 179; Mismatches 197; Indels 6; Gaps 5;  
QY 1 MGPTSGPSSL--LILLTHLPLALGSPWYSITIPNIRLESEETMVLAEADAGQGVFVTVT 58  
DB 1 MGPASGQQLVLLALLASSPLALGIPWYSITIPNVLRESSETIVLEAHDAGQDIPVTVT 60  
QY 59 VHDFFGKVLJLSSEKTVLTATNMGNVTFTPANREFKSEKGRNKFVTVQATFGTQVE 118  
DB 61 VQDFL-KRQVLTSEKTVLTGASGHLRSVSIKIPASKEFNSDEKGHKYVTVVANFGETVVE 119  
QY 119 KWLVSLSQSGYLFQTDTKTYTPGSTVLYRIFTVNHKLLPVGRTVMNIENPEGIPVKQD 178  
DB 120 KAVWVFQSGYLFQTDKTYTPGSTVLYRIFTVNDNLLPVGKTVVILLIETPDGIPVKRD 179  
QY 179 SLSSQNLGVLPLSWDIPELVWNGQWKIRAYENSPOQVSTEFVEKVEYVLPSEVIVEP 238  
DB 180 ILSSNQHGLPLSNIPELVWNGQWKIRAFYEHAPKQIFSAEFVEKVEYVLPSEVIVEP 239  
QY 239 TEKPYIYNKGLVITITAFLYGKVEGTAFVFIQDGEQRIPLPSLKRIFPIEDSG 298  
DB 240 TETFYIDDPNGLEVSIIAKFLYKKNVDTAFVFIQDGEQRIPLPSLKRIFPIEDSG 299  
QY 299 EWLRSKVLDDGQNFPAEDLVGKSLVVSATVLLHSGSDMVQAESGPIVTSVQVHFT 358  
DB 300 DAVLRKVLMEGVAPSNADALVGKSLVVSATVLLHSGSDMVQAESGPIVTSVQVHFT 359  
QY 359 KTPKYFPGMPFDLMVFTVNPDGSPAYRVFVAVQGEDTVQSLTQDGVAKLSINTHPSQK 418  
DB 360 KTPKFFKAMPFDLMVFTVNPDGSPASKVLVTVTQGSN-AKALTQDQGVAKLSINTHPSRQ 418  
QY 419 PLSTVTKKQELSEAEQARTVQALPYSTVGNNSNLYLHSLVRLTELPGETLVNVELR 478  
DB 419 PLTTVTTKDTPESQATKWEAHPISTMGNSNLYLHSLVRLTELPGETLVNVELR 478  
QY 479 MDRAHEAKIRYYTYLIMNKGKLLKAGQVREPGQDLWLPISITTDIFPSRLVAYTYLI 538  
DB 479 TDPGHEAKIRYYTYLIMNKGKLLKAGQVREPGQDLWLPISITTDIFPSRLVAYTYLI 538  
QY 539 GASQREWVADSVWVDKSCVSLVVKSGQSEDRQVPGQQTMLKIEGDHARVWLVAV 598  
DB 539 GASQREWVADSVWVDKSCVSLVVKSGQSEDRQVPGQQTMLKIEGDHARVWLVAV 597  
QY 599 DKGVFLVKNKLTQSKINDVVEKADIGCTPSGKDYAGVPSDAGLTITSSSGQQTQARA 658  
DB 598 DKGVFLVKNKLTQSKINDVVEKADIGCTPSGKDYAGVPSDAGLTITSSSGQQTQARA 657  
QY 659 ELQCPQFAARRRSVQLTERMDKVGKY-PKSLRKCCEDGMEENPMWRSQRRTRFISLG 717  
DB 658 DLECTKPAARRRSVQLMERMDKAGQYTDKGLKCCEDGMDIPWYSCQRRARLIQ 717  
QY 718 EACKVFLDCNVTITELARQHAASHGLARSLNLEDDIIEENIVSRSEFPESMWLVYED 777  
DB 718 EACKVFLDCNVTITELARQHAASHGLARSLNLEDDIIEENIVSRSEFPESMWLVYED 777

Db 718 ENC1KAFIDCNHITKURBQHRDHLVGLARSELBEDIIPBEDIISRSHPFOSWLTWIE 777  
QY 778 LKPPKRGISTKLNFIKDSITTWELAVSMGDKGICVADFEVTVNQDFFIDRLRPY 837  
Db 778 LKPPKRGISTKLNFIKDSITTWELAVSMGDKGICVADFEVTVNQDFFIDRLRPY 837  
QY 838 SVRNEQVEIRAVLYNRQCEKLVKRVVLELHNAFGLATTKERHOOTITIPKSSLSVP 897  
Db 838 SVRNEQVEIRAVLYNRQCEKLVKRVVLELHNAFGLATTKERHOOTITIPKSSLSVP 897  
QY 898 YVIVPLKTGQEVVKAAYVHFTISDGRKSLKVPKRGIMNKTVAVRTLDPERLREGV 957  
Db 898 YVIVPLKTGQEVVKAAYVHFTISDGRKSLKVPKRGIMNKTVAVRTLDPERLREGV 957  
QY 958 QKEDIPADISDQVDPDTSESTRILLQGTPTVAQMTEDAVDAERLKHILVTPSCGQGNMIG 1017  
Db 958 QKEDIPADISDQVDPDTSESTRILLQGTPTVAQMTEDAVDAERLKHILVTPSCGQGNMIG 1017  
QY 1018 MPTVIAVHYLDETEQWEKFGLEKROGALBELIKKGYTOQLAFQPSAFAAFVKAAPSTW 1077  
Db 1018 MPTVIAVHYLDETEQWEKFGLEKROGALBELIKKGYTOQLAFQPSAFAAFVKAAPSTW 1077  
QY 1078 LTAYVVKVFLAVNLIAIDSVOLCGAVKWLILEKQKPDGVFOEDAPVHOEIMIGLRNN 1137  
Db 1078 LTAYVVKVFLAVNLIAIDSVOLCGAVKWLILEKQKPDGVFOEDAPVHOEIMIGLRNN 1137  
QY 1138 EKDWALTAFLVLSIOEAKDICEQVNSLPGISITKAGDFLEANNYMLQRYTVIAIAGYALA 1197  
Db 1138 EADVSLTAFLVLSIOEAKDICEQVNSLPGISITKAGDFLEANNYMLQRYTVIAIAGYALA 1197  
QY 1198 QMGLKPLNKLFTTAKDKNRVEDPGQKLYNVEATSYALLALLKQDFDPPVVRWLN 1257  
Db 1198 QMGLKPLNKLFTTAKDKNRVEDPGQKLYNVEATSYALLALLKQDFDPPVVRWLN 1257  
QY 1258 BORYGGYGSTQATFWFOALAQYQKADPHOELNLDVSLQIPRSKITHRIHWSAS 1317  
Db 1258 BORYGGYGSTQATFWFOALAQYQKADPHOELNLDVSLQIPRSKITHRIHWSAS 1317  
QY 1318 LLRSEETKENEFGVTVAEGKGQGTLSVVTMYHAKADQLTCKNFKDLKVTIKAPATEKRP 1377  
Db 1318 LLRSEETKENEFGVTVAEGKGQGTLSVVTMYHAKADQLTCKNFKDLKVTIKAPATEKRP 1377  
QY 1378 QDAKNTMLEICHTYRGDQATWSILDSMTGAFAPDQDLDQLANGVDVRIKVELDKA 1437  
Db 1378 QDAKNTMLEICHTYRGDQATWSILDSMTGAFAPDQDLDQLANGVDVRIKVELDKA 1437  
QY 1438 FSNKNTLIYLEKISHEEDCLTFKVHQYFNVGLIQGSKVYVYNNLEESCTRFYHPEK 1497  
Db 1438 FSNKNTLIYLEKISHEEDCLTFKVHQYFNVGLIQGSKVYVYNNLEESCTRFYHPEK 1497  
QY 1498 EDGLNKLCRDELCEACENCFQKSDDKVTLERLDKACERGVYVYKTRLVKVLQNSD 1557  
Db 1498 EDGLNKLCRDELCEACENCFQKSDDKVTLERLDKACERGVYVYKTRLVKVLQNSD 1557  
QY 1558 FDEYIMAEQITKSGSDEVQVQORTFISPIKREALKLEKXKHYLMWGLSDFWGEKPN 1617  
Db 1558 FDEYIMAEQITKSGSDEVQVQORTFISPIKREALKLEKXKHYLMWGLSDFWGEKPN 1617  
QY 1618 LSYIIGKDTVHHKPEDEQDQENQKQCDLCAFTSMVVGCPN 1663  
Db 1618 LSYIIGKDTVHHKPEDEQDQENQKQCDLCAFTSMVVGCPN 1663

RESULT 7  
Q9GKPL  
ID Q9GKPL PRELIMINARY; PRT; 1661 AA.  
AC Q9GKPL  
DT 01-MAR-2001 (TremBurel. 16, Created)  
DT 01-MAR-2001 (TremBurel. 16, Last sequence update)  
DT 05-JUL-2004 (TremBurel. 27, Last annotation update)  
DE Complement component C3 (Complement C3).  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=21313047; PubMed=11419349;  
RA Wimmers K., Mekchay S., Ponsuksili S., Hardge T., Yerle M.,  
Schellander K.;  
SA "Polymorphic sites in exon 15 and 30 of the porcine C3 gene";  
RT Anim. Genet. 32:46-47(2001).  
RL [2]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Wimmers K., Ponsuksili S., Schmol F., Schellander K.;  
RL Submitted (JUL-2002) to the ENBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=22444329; PubMed=13557058;  
RA Wimmers K., Mekchay S., Schellander K., Ponsuksili S.;  
RT "Molecular characterization of the pig C3 gene and its association  
with complement activity";  
RL Immunogenetics 54:714-724(2003).  
DR ENBL; AP154933; AAG40565.1; -;  
DR ENBL; A494748; CAD38823.2; -;  
DR HSP; P01026; IQP.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0004866; F:endorpeptidase inhibitor activity; IEA.  
DR GO; GO:0006956; P:complement activation; IEA.  
DR GO; GO:0006954; P:inflammatory response; IEA.  
DR InterPro; IPR002890; A2M N.  
DR InterPro; IPR009048; A2M N.  
DR InterPro; IPR000020; Anaphylatoxin.  
DR InterPro; IPR001840; Anaphylatoxin.  
DR InterPro; IPR008964; Invasin.  
DR InterPro; IPR001599; Macroglobulin.  
DR InterPro; IPR001134; Netrin C.  
DR InterPro; IPR008930; Terc cyc. toroid.  
DR InterPro; IPR008993; TIMP-like.  
DR Pfam; PF00207; A2M; 1.  
DR Pfam; PF01835; A2M N; 1.  
DR Pfam; PF01821; ANATO; 1.  
DR Pfam; PF01759; NTR; 1.  
DR PRINTS; PR00004; ANAPHYLATOXN.  
DR ProDom; PD003264; Anaphylatoxin; 1.  
DR SMART; SMC0104; ANATO; 1.  
DR SMART; SMC0643; C345C; 1.  
DR PROSITE; PS00477; ALPHA 2 MACROGLOBULIN; 1.  
DR PROSITE; PS01177; ANAPHYLATOXIN 1; 1.  
DR PROSITE; PS01178; ANAPHYLATOXIN 2; 1.  
DR PROSITE; PS0189; NTR; 1.  
SQ SEQUENCE 1661 AA; 186805 MW; 4899D0914BE3310C CRC64;

Query Match 78.1%; Score 6724; DB 2; Length 1661;  
Best Local Similarity 76.5%; Pred. No. 0;  
Matches 1272; Conservative 189; Mismatches 200; Indels 2; Gaps 2;

QY 1 MGPTSPSLLLLHPLALGSPMYSLITNLRLESEEMVLEAHDAGDYPVTVH 60  
Db 1 MGSTSGPRLLLLLSLPALGDPPIYITITNVRLESEEMVLEAHDAGDYPVTVH 60  
QY 61 DFFGKLVLSSEKTVLTATNMGNVTFTIPANREFKSEKGRNKFVTVQATFGTVVEK 120  
Db 61 DFFAKQVLSSESTTTNNANNVLTSTNLIKIPASKEFKSEKQ-HKFTVTVQALFGNVQVEK 119  
QY 121 VLVSQSGVLFOTDKTIYTSQVLEIFVTHNKLPLVGTVMVNTLENPEGIPVKDLS 180  
Db 120 VLVSQSGVLFOTDKTIYTSQVLEIFVTHNKLPLVGTVMVNTLENPEGIPVKDLS 179  
QY 181 SSQNLGVLPLSWDIPELVNMGMKIRAYNSPQVSTFEVKEVYVLPSPFIVEPTE 240  
Db 180 SSNQPGILALSWNIPELVNMGMKIRAYNSPQVSTFEVKEVYVLPSPFIVEPTE 239

QY 241 KFYIYNEKLEVTITARELYKKVEGTAFVIFGIQDGEORISLSPESIKRIPIEDSGSEV 300  
DB 240 KFYIYDIDPGLTUNIIARELYGESVDGTAFAVIFGVQDQDQISLSQSLTRVPIIDGTGEA 299  
QY 301 VLSKVLGQVQNPRAEDLVGSLVYSATVILHSGSDMVQASRGIPVTSPIYHFTKT 360  
DB 300 TLOGVLLNGVHSSVNDLVGSIYVSATVILHSGSDMVQASRGIPVTSPIYHFTKT 359  
QY 361 PKYFKGMPDLVVFVNDPDSPAYRVVAVGEBTVQSLTQGDGVAKLSINTHPSQKPL 420  
DB 360 PKYFKGMPDLVVFVNDPDSPAYRVVAVGEBTVQSLTQGDGVAKLSINTHPSQKPL 418  
QY 421 SITVTRKQELSAEQATRMQALPYSTVGNENMNYLHLSVLRTLPGETLNVNFLRMD 480  
DB 419 PITVTRKQELSAEQATRMQALPYSTVGNENMNYLHLSVLRTLPGETLNVNFLRMD 478  
QY 481 RAHEAKRYTYTILIMKGRLLKAGROVREPQGLVVLPLSTITTFIPSRFLVAYTILGA 540  
DB 479 PGYQDKIRYFTYILIMKGRLLKAGROVREPQGLVVLPLSTITTFIPSRFLVAYTILGA 538  
QY 541 SGOREVADVVDVNDKSDCVGSLVKSQSEDRQVPGQOMTLKTEGHDGARVVLVAVDK 600  
DB 539 NGOREVADVVDVNDKSDCVGSLVKSQSEDRQVPGQOMTLKTEGHDGARVVLVAVDK 598  
QY 601 GVFVNLKXKLTOSKIDWVVEKADIGCTPGSGKDVAGVPSDAGLTFTSSSQOQTAQABE 660  
DB 599 GVFVNLKXKLTORRWDVVEKADIGCTPGSGKDVAGVPSDAGLTFTSSSQOQTAQABE 658  
QY 661 CQPOAARRRSVQLTEKMDKVGKYPKELRCCEDGRENPMRPFSCORRTFTSLGEAC 720  
DB 659 ECPKFAARKERSVQLMEKMDKLGQYKDVRCCEHGMEDNPMKESQORRAQFTLHGDCAC 718  
QY 721 KVFLLDCNVITLRRQHARASHLGLARSNLEDIIARENVISSEPSWLNWVELDK 780  
DB 719 VKAPLDCCEVIAKROHGRNPLGLARSNLEDIIARENVISSEPSWLNWVELDK 778  
QY 781 PPKNGISTKLMNIFLKDSITTWELLAVNSDKKGCIVADPFVTVMDQDFIDRLPLSVV 840  
DB 779 PPKNGISTKLMNIFLKDSITTWELLAVNSDKKGCIVADPFVTVMDQDFIDRLPLSVV 838  
QY 841 RNEQVEIRAVLYNRQBELKVRVLLHNPACSLATTKRHHOQTIIPKSSLSVPIYI 900  
DB 839 RNEQVEIRAVLYNRQBELKVRVLLHNPACSLATTKRHHOQTIIPKSSLSVPIYI 898  
QY 901 VPLKTLQEVVEKAAVYHFIISGVKRSKLVVPEGRMKNKTVAVTLLDPERLREGVQKE 960  
DB 899 VPLKTLQEVVEKAAVYHFIISGVKRSKLVVPEGRMKNKTVAVTLLDPERLREGVQKE 958  
QY 961 DIPADLSQVPTSETRILQGTVEVQMTEDAVDABRLKHLIVTPSGGGEQNNIGMT 1020  
DB 959 DIPADLSQVPTSETRILQGTVEVQMTEDAVDABRLKHLIVTPSGGGEQNNIGMT 1018  
QY 1021 TVIAVHYLDTEQWFKLEKQGALELIKGYTQQLAPRQPSAFAPVVKRAPSSTWLT 1080  
DB 1019 TVIAVHYLDTEQWFKLEKQGALELIKGYTQQLAPRQPSAFAPVVKRAPSSTWLT 1078  
QY 1081 YVVKVFLAVNLIAIDSVLCGAVKWLILEKQKPGVFCEDAPVTHOEMIGELNNNEKD 1140  
DB 1079 YVVKVFLAVNLIAIDSVLCGAVKWLILEKQKPGVFCEDAPVTHOEMIGELNNNEKD 1138  
QY 1141 MALTAFLVLSIQBAKIDCEQVNSLPGSTIKAGDFLEANYNLSQSYTVAIAGVALQMG 1200  
DB 1139 VSLTAFVLLAQBAKIDCEQVNSLPGSTIKAGDFLEANYNLSQSYTVAIAGVALQMG 1198  
QY 1201 RLKGPLNLKFLTTAKDNWEPKQVLYNVEATSVALLALQLKDFDFVPPVVRWLNQ 1260  
DB 1199 KLDEPFLNLKFLTTAKDNWEPKQVLYNVEATSVALLALQLKDFDFVPPVVRWLNQ 1258  
QY 1261 YGGGYSQSTQATPWFQALAOYKQADPHQELNLSQSLPSRSKITHRHINESASLLR 1320  
DB 1259 YGGGYSQSTQATPWFQALAOYKQADPHQELNLSQSLPSRSKITHRHINESASLLR 1318  
QY 1321 SEETKENEGFTVTAEGKGQGTLSVVTMYHAKAKDLCTCNKFDLVKTVKPAFETEKRPQDA 1380

DB 1319 SEETKENEGFTVTAEGKGQGTLSVVTMYHAKAKGTCTCKFKLVSIHPAPEPVKKQPEA 1378  
QY 1381 KNTWILICITRYRGDOOATMSILDIISMTGFPADPTDQLKQLANGVDVRIKVELDKAFSD 1440  
DB 1379 KSSNVLDICITRYLGNDQATMSILDIISMTGFPDIEDLKLSTGVDRVYSKVELNALS 1438  
QY 1441 RNTUILLVKSSEHDDCLAFKQVQFNVVELIQPGAVKVIYVYNNLEBCTRFYHPEKEDG 1500  
DB 1439 KNTLIIYLDKISHTLEDICISFKVQFNVVGLIQPGSKVYSYINLDECTRFYHPEKEDG 1498  
QY 1501 KNLKLCDELCRCAEENCFQKSDDKVTLERLDCACEPGVDVYVYKTLVKVQLSNDPDE 1560  
DB 1499 MLNKLCHKRCRCAEENCFPMHDEEVTLLDRLERACEPGVDVYVYKTLVKVQLSNDPDE 1558  
QY 1561 YVMAIETIISGSDVQVQGTFFISPIKCRALKLBEKKHYLMGLSDDFWGKPNLSY 1620  
DB 1559 YVMAIETIISGSDVQVQGTFFISPIKCRALKLBEKKHYLMGLSDDFWGKPNLSY 1618  
QY 1621 IIGKDTVVEHWPBEDECCQDEENKQCDLCAFTESVWVFGCPN 1663  
DB 1619 IIGKDTVVEHWPBEDECCQDEENKQCDLCAFTESVWVFGCPN 1661

RESULT 8  
Q90633 PRELIMINARY; PRT; 1652 AA.  
AC Q90633;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Complement C3 precursor.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
NCBI\_TaxID=9031;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver.  
RX MEDLINE=95173423; PubMed=7532662;  
RA Mavroidis M., Sunyer J.O., Lambiris J.D.;  
RT "Isolation, primary structure, and evolution of the third component of  
RT chicken complement and evidence for a new member of the alpha 2-  
RT macroglobulin family.",  
RL J. Immunol. 154:2164-2174 (1995).  
DR EMBL; U16848; AAA64694.1; -.  
DR PIR; I50711; I50711.  
DR HSP; P01024; LGHO.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0004866; F:endorpeptidase inhibitor activity; IEA.  
DR GO; GO:0006956; P:complement activation; IEA.  
DR GO; GO:0006954; P:inflammatory response; IEA.  
DR InterPro; IPR002890; A2M\_N.  
DR InterPro; IPR009048; AM\_receptor\_bind.  
DR InterPro; IPR001840; Anaphylatoxin.  
DR InterPro; IPR001598; MacroglobinA2.  
DR InterPro; IPR001134; Netrin\_C\_toroid.  
DR InterPro; IPR008930; Terp\_cyc\_toroid.  
DR Pfam; PF00207; A2M; 1.  
DR Pfam; PF01835; A2M\_N; 1.  
DR Pfam; PF01821; ANATO; 1.  
DR Pfam; PF01759; NTR; 1.  
DR PRINTS; PR00004; ANAPHYLATOXN.  
DR ProDom; PD003264; Anaphylatoxin; 1.  
DR SMART; SM00104; ANATO; 1.  
DR SMART; SM00643; C345C; 1.  
DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.  
DR PROSITE; PS01177; ANAPHYLATOXIN\_1; UNKNOWN\_1.  
DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.  
DR PROSITE; PS0189; NTR; 1.

KW Signal.  
 FT CHAIN 1 16 Potential.  
 FT CHAIN 644 1652 Complement C3 alpha chain.  
 FT CHAIN 17 639 Complement C3 beta chain.  
 SQ SEQUENCE 1652 AA; 18408 MW; 9020C49FE127BD1F CRC64;

Query Match  
 Best Local Similarity 52.5%; Score 4518.5; DB 2; Length 1652;  
 Matches 896; Conservative 305; Mismatches 435; Indels 41; Gaps 24;

QY 1 MGFTSGPGLILLTHPLALGSP--MYSIITENILRLSEETMVLNHAQDQVPTVT 58  
 DB 1 MGLLLPLLLGLVLLH--AVPTPAQMTVMTPAVRLDLDKRVLEAPGLSAPTEANLL 57  
 QY 59 VHDFFGKLVLSSEKTLVTPATNMGVNTFTIPANREFKSEKGRNKFVTVQATFGQVVE 118  
 DB 58 VQDFPQKRVLFQVKQLNPAEGMMAIATVKVPV-KLLPPVKGK-FVSVVARVQGVTL 115  
 QY 119 KVLVLSQGLVLTOTOKTIVTPGSTVLYRIFTVNHKLLPVGTVMVNIENEGIPKOD 178  
 DB 116 KVLVLSQGLHIFQTDKPIVTPGSTVLSRLFALSHFMQPLLKTVIVEKVTNDNVIKOV 175  
 QY 179 SLSSQNLGLVPLSDIPELYNMQWKIRAYENSPQVFEFTEFVEKYLVSFVIVEP 238  
 DB 176 PVSSPMRNGIFSINHLPEVUSLGTWITAKFDSQVQVFEFTEFVEKYLVSFVITLDP 235  
 QY 239 TEKFYIYNKGLVITITARLYGKVEGTAFLVIFGIDGEOEISIPESKRIPIEDGSG 298  
 DB 236 QKFYIIPAEFRVITITARLYGKNLQGTAFVLFGVVDEKKTIPQSLQKVKTDDG 295  
 QY 299 EVLRSKVLGQVQNPRAEDLVGKSLVSATVILHSGSDMVQAESGIPVTSPIQIHT 358  
 DB 296 QAVLPWMLRQPPAN--LQELVGHSLVTVTVLTESGSDMVQAESGIRIVTSPIYHT 353  
 QY 359 KTKPKFKGMPFDLWFTVNPDSGPAYR-VPVAVQGEDTVQSL--TQDGVKALISINTP 415  
 DB 354 HTPKFKGMPFDVTVVNTDNPSAAGIPVKA--DNFQGLVSTQRTGAKVLNMPA 410  
 QY 416 SKPLSITVTKKQELSAEQAATMOALPYSTVGNNSNYLHLSVRLTELPGETLNVNF 475  
 DB 411 NKNSPFTVTRTDQDLPPEROASQIVAEAYQSGNSGNYLHLAVASQVQPDNLPINF 470  
 QY 476 LURMDRAHEAK-IRYTYLIMNKGKLLKAGQVREPODLVPLSLITDFTPSRLVAY 534  
 DB 471 HLKSNRDDVRKSVSYFTYLLSKGHIHVHVGQREGQSLVTSVPTANLPSFRVAY 530  
 QY 535 YTLIGASQREVAVSWVDVKSCVGLVVK-SQSEDRQPVPGQOMTLKTEGDHGRV 593  
 DB 531 YHV---RPGELIADSVWVDKTCMGLSVVRGASEADNRVHPRTPMELHIEGDHKA 586  
 QY 594 VLVAVDKGVFLNKKNTQSKIMDVVEKADIGCTPGSKGDYAGVFSAGLFTSSGQ 653  
 DB 587 GLVAVDKAVYVLN-KNKLTSQKNDTVENSIDGCTPGSGRNOVGVFADAGLSLTSNVNIN 645  
 QY 654 TAQRAELQCPQAPARRRSVQLTEKMDKUGKY-PKELRKCCDGHRENPMRPSCORRTR 712  
 DB 646 TEQSEVQCAKPAKGRKRSVRLIKHGTQMAEYSDKNLRKCCDGGIRKMLMGVSCBKAT 705  
 QY 713 FTSLSGACKVFLDCNVTITBLRQCHARASHLGLARSNLDEDIIASENIVSRSEPPESML 772  
 DB 706 VYLDKASTCEAPLSCLYKIGIRDEE-RELQVELARSEVDDAFLSDSDITSRLSFPESWL 764  
 QY 773 WNVDELKBPB-KNGISTKLMIIFLKDSITTWELIAMSDDKIGICVADPEFVTVMQDPTI 831  
 DB 765 WQVEELTBPPEQGSMTLPIYLDKDSITTWELIAMSSENKGLCVADPEYIITWKEFFI 824  
 QY 832 DLRLPYSVVRNEQVEIRAVLYNRQNLKVRVRELLNPAFCSLATTKRRHQQTITIPPK 891  
 DB 825 DLRLPYSVAVRNEQVEIRAILNYWTN-KIKVRVELMYPALCSASTTKTRYQYIQLPEBQ 883  
 QY 892 SLSVPEYIVPLKGLQVEVEKAAVYHFIISDGVKSLKVVPEGIRMNKTVARTIDPER 951  
 DB 884 SSDAVFVPLVLELQGHQVHVEKAAVWNSFVSDGVKKKRLRVVPSGMRLEKTVKVELDPKT 943

QY 952 LGREGVOKEDIPADLSDOVPDTESETRILLQCTPVAQWTEADVAERLKHILVTPSGCG 1011  
 DB 944 LGNNGVQEVKVKAAMLSDIVPNTSETKVSIQGNPVSILVEKATDGTGKHLIVTPSGCG 1003  
 QY 1012 EQNMIGMTPTVIAVHYLBETEWEKXFGLEKQGALELIKKGYTQQLAPROPSSFAAFVK 1071  
 DB 1004 EQNMIGMTPTVIAVHYLBETEWEKXFGLEKQGALELIKKGYTQQLAPROPSSFAAFVK 1063  
 QY 1072 RAPTWTWIAVVKVSLAVNLIAIDSOVLGAVKWLILEKOKPGVQEDAPVTHQWIMIG 1131  
 DB 1064 RPTSTWIAVAVKVPAMAINVVDIAPFVVCAGIKWILILEKQOPDGLFOEDAPVHKEMVG 1123  
 QY 1132 GLANNEKDMALTAFVLISLOPAKDI CEQVNSLPGSTTKAGDGFLEANYMQLQSYVAI 1191  
 DB 1124 GY-HGAEPSVSLTAFVLSALQESQKIKNYVKSLDGSIAKSDYLSRKYOSLTPYTV 1182  
 QY 1192 AGVALAQMGRKLG-PLLNKFLTTAKDKRWBEPGKQLNVEATSVALLALLQLKDPDFVP 1250  
 DB 1183 TSVALALTGLKSEKVMKF---SKDGTWAEARNATVNIESTYALVALLOMEKALTG 1239  
 QY 1251 PVYRWLNBOYRGYGGYSTQATFMVFOALAOYQKADPDHQLNLDVSLQPSRSKITHR 1310  
 DB 1240 PVYRWLAQNYGGYSTQATILFVFOALAOYHVALPEHVELNLDVSVLLPRANATYR 1299  
 QY 1311 IHVESALLRSBETKENEGFTVTAEGKQGTLSVVVMTYHAKAD-QLTCNKFDLKV 1366  
 DB 1300 IENNALVARSATKLNEDFTVKAEGTKGIMTVVTVVYKAVPEKENKCNDFLRVSVED 1359  
 QY 1367 IKAPATEKRPQADKMTMLEICTRYRGDDATWSILDISMGTGAPDITDILKOLANGVD 1426  
 DB 1360 VKAGREVEGVIRSVKIT---ICTRFLDVTDATWSILDISMGTGAPDITDILKOLANGVD 1415  
 QY 1427 RYISKVELKAPSDRNTLIIYLDKVSHEDEDCIAPKHYFNVELIQGAVKVVAYNLE 1486  
 DB 1416 RYISKFEIDHALSNRNLIIYLDKVSHEDEDCIAPKHYFNVELIQGAVKVVAYNLE 1475  
 QY 1487 ESCRTHYHPEKSDGKLNKLCRDELRCABENCFIQ-KSDDKVTLEERLDKACEPQVDVY 1545  
 DB 1476 DRCTRYPHDPKAGGQURKICHGEVC-CABENCFIKVKNDNPITVNERIDLACKPGVDVY 1534  
 QY 1546 KTRLVKQVQNSNDPDEYMAIEQTIKSGDVEVOGQORTFISPIKREALKLEKKHYLMW 1605  
 DB 1535 KVKVATEETPPSHNDYMSILTVIKVGTDPENPGSNRIFVSHKQCRDALSLOKQDYLW 1594  
 QY 1606 GLSDDPWGKPNLSYIIGKQVWEHVEPREDECEQKQCDLGAFTESMVFGCP 1662  
 DB 1595 GLASDLVWTCGRSPSYLISKDTWLEAWPLEESQDADLQPLCQDFTFESDNLVLFQCP 1651

RESULT 9  
 CO3\_NAJNA  
 ID\_CO3\_NAJNA STANDARD; PRT; 1651 AA.  
 AC Q01533;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Complement C3 precursor [Contains: C3a anaphylatoxin].  
 GN Name=C3;  
 OS Naja naja (Indian cobra).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroides;  
 CC Ophiidia; Elapidae; Naja.  
 OX NCBI\_Taxid=35670;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=3056528; PubMed=1431125;  
 RA Fritzinger D.C., Connolly M., Petrella E.C., Bredehorst R.,  
 RA Vogel C.W.;  
 RT "Primary structure of cobra complement component C3";  
 RL J. Immunol. 149:3554-3562(1992).  
 CC !- FUNCTION: C3 plays a central role in the activation of the

complement system. Its processing by C3 convertase is the central reaction in both classical and alternative complement pathways. After activation C3b can bind covalently, via its reactive thioester, to cell surface carbohydrates or immune aggregates. C3a anaphylatoxin is a mediator of local inflammatory process. It induces the contraction of smooth muscle, increases vascular permeability and causes histamine release from mast cells and basophilic leukocytes.

-1- SUBUNIT: C3 precursor is first processed by the removal of 4 Arg residues, forming two chains, beta and alpha, linked by a disulfide bond. C3 convertase activates C3 by cleaving the alpha chain, releasing C3a anaphylatoxin and generating C3b (beta chain + alpha chain) (By similarity).

-1- SIMILARITY: Contains 1 anaphylatoxin-like domain.

-1- SIMILARITY: Contains 1 NTR domain.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; L02365; AAA49385.1; -

HSP; P01026; 1Q5J.

InterPro; IPR002890; A2M N.

InterPro; IPR009048; AM\_Receptor bind.

InterPro; IPR000020; Anaphylatoxin.

InterPro; IPR001840; Anaphylatoxin.

InterPro; IPR008964; Invasin intimin.

InterPro; IPR001599; MacroglobinA2.

InterPro; IPR001134; Netrin C.

InterPro; IPR008930; Terp cyc toroid.

InterPro; IPR008993; TIMP\_like.

Pfam; PF00207; A2M; 1.

Pfam; PF01835; A2M\_N; 1.

Pfam; PF01821; ANATO; 1.

Pfam; PF01759; NTR; 1.

PRINTS; PR00004; ANAPHYLATOXIN.

ProDom; PD003264; Anaphylatoxin; 1.

ProSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.

ProSITE; PS01177; ANAPHYLATOXIN\_1; 1.

ProSITE; PS01178; ANAPHYLATOXIN\_2; 1.

ProSITE; PS0189; NTR; 1.

Complement alternate pathway; Complement pathway; Inflammatory response; Plasma; Signal; Thioester bond.

SIGNAL 1 22

CHAIN 23 1651 Complement C3.

CHAIN 23 655 Complement C3 beta chain.

CHAIN 661 1651 Complement C3 alpha chain.

PEPTIDE 661 738 C3a anaphylatoxin.

CHAIN 739 1651 Complement C3b alpha' chain.

CHAIN 683 718 Anaphylatoxin-like.

DONAIN 1506 1649 NTR.

SITE 738 739 Cleavage (by C3 convertase).

DISULFID 546 807 Interchain (By similarity).

DISULFID 615 650 By similarity.

DISULFID 683 710 By similarity.

DISULFID 684 717 By similarity.

DISULFID 697 718 By similarity.

DISULFID 863 1501 By similarity.

DISULFID 1091 1477 By similarity.

DISULFID 1477 1446 By similarity.

DISULFID 1494 1499 By similarity.

DISULFID 1506 1578 By similarity.

DISULFID 1525 1649 By similarity.

CROSSLNK 999 1002 Isoglutamyl cysteine thioester (Cys-Gln) (By similarity).

SEQUENCE 1651 AA; 184926 MW; 66F91BD8F4A935F7 CRC64;

Query Match 51.0%; Score 4389.5; DB 1; Length 1651;  
Best Local Similarity 51.3%; Pred. No. 1.1e-252;  
Matches 853; Conservative 309; Mismatches 471; Indels 31; Gaps 18;

QY 9 LLLALLHLPLALGSPMYSIIITENILRLSEEMVLEAHDAQDVPVTVTHDFPGKVL 68  
DB 8 LVAALLGFGSSHGALYLIITPAVLTTEBOILVEAGDSTPKSLDIFVHDFPRKQT 67  
QY 69 LSSEKTVLTTPATHNMGNVFTTIPANREFKSEKGRNKFVTVQATFTQV-VEKVLVLSQS 127  
DB 68 LFQSRVDMNAGSMFVPTTIKVA-KELNKDSKQNYVYVKT-GFOVALEKWLSSQS 125  
QY 128 GYLFIQTDKTIYTPGSLVYRIETVNHKLLPVGRVTMNIENPEGIPVKQDSLSSQNLG 187  
DB 126 GFVFIQTDKGLIYTPGSPVRVRSVDHNRDKTIVFEFQTEGILVSSKPVNPSGSI- 184  
QY 188 VLPSWDPIDELVNMGMKIRAYIENSPOQVFSTFEVKEVYLPSPSEVIEPTEKYYIYN 247  
DB 185 ---RPYNLPVLVSGTGWAKAYEHSPESTAYFVREYVLPSPSEVRLQPSDKELYIDG 241  
QY 248 EKGLEVTITARFLYGGKVEGTAFVIFGIODGEORISLPESLKRIPTEDSGSEWLSRVL 307  
DB 242 KKNFHVITARYLYGKKEGVAVFVGKIDDAKIPDSLTRIPIIDGCBATLKROTLL 301  
QY 308 LDGVNPRABDLVKSILYVSATVILHSGSDMVAERSGPIVTSPPYQIHFYKPYKPG 367  
DB 302 RSRFQD-LNQLVGHVTLVSVTVITBSGSDMVVTEQGGIHTVTSYQIYFTKPYKPG 359  
QY 368 MPFDLMVFTNPDGSPAYRVPV--AVQGEDTVQSITQDGVAKLSINTHPQKPLSITV 424  
DB 360 MPYELTVVTPDGSRAHVVPVVSSEAHSEGT---TLDGTAKILNLTPLNIQSLPITV 415  
QY 425 RTKQOEISEAEQATRNQALPYSTVNGSNYHLVRLTELPGETLVNFWLLRMDRAHE 484  
DB 416 RTNHDGLPRERQAIKSMATATAYQTQGSSENYLHVATISTEIPGDNLPFNFNVRGNANSL 475  
QY 485 AKIRYTYVLMNKGRLKAGRVREPGQDLVLPISITTDPIPSRLVAYVYTLIGASQR 544  
DB 476 NQIKFYVILINKGKIFKVGQPRRQGNLVNWLHITPDLFSPRFVAYVQV---GNN 531  
QY 545 EVVADSVWVDKSCVGLVWKSQS-EDRPVPGQOMTLKIEGDHARVLVAVDKGVF 603  
DB 532 EIVADSVWVDKTCMGTCLVVGASSRDDRIRQPGAAKIKLEGDPGARVGVAVDKAVY 591  
QY 604 VLNKNKLTQSKWDVVEKADIGCTPGSKDYGACVFSADAGLTFTSSSGOOTQRAELQCP 663  
DB 592 VLNDKYKISQAKINDTIEKSDFGCTAGSGQNNLVGFEDAGLALTSTNLNTKQSAKCP 651  
QY 664 QPAARRRR-SVOLTEKMDKVGKYPKE-LNKCCEDGMRENMRFSQRTTRISLGEACK 721  
DB 652 QPANRRRRSSVLLDLSKASAAQFQDGLKCCEDGMHENPMGYTCEKAKYIQGDACK 711  
QY 722 KVFLDCNNYITELRRQSHARASHGLARSNLDEIIAENIVSRSEPPESLWNVNEDLK-E 780  
DB 712 AAFLECCHYIKGIEDENQRESEIFLAESDFEDELFGDDNIISRSDPFESWMLTTELTCGE 771  
QY 781 PPQNGISTKLMNIFLKDSITTWILLAVSMCKGICVADPFVTVQMDFFIDLRIPYSVY 840  
DB 772 FNNQGISSTKVPFYLRDSITTWELLAVGLSPTKGCIVABFEYITWMDKDFIDLRIPYSV 831  
QY 841 RNEQVEIRAVLYNRQNELKVRVELLHNPAFSLATTKRRHQQTITIPKSSLSVYVI 900  
DB 832 KNEQVEIRAILYNY-ADEDIVRVVELLYNPAFCASTEGRYVQQFPPIKALSSRAVPFI 890  
QY 901 VPLTKGLOEVVEKAAVTHHFISQVRKSLKVPVEGRMKTKVAVRTLDERLGRGVQKE 960  
DB 891 VPEQGLHDVEVIASVARGELASDGVKRLKVPVEGERKN-IVTIIELDPSVGVGTQEL 949  
QY 961 DIPPADLSQDQVPTDESETRILLQGTTPVAQMTEDAVDAERLKHILVTFSCGSGQNIGMTP 1020  
DB 950 TVIANKLDDKVPDTEVEITRISVLGDPAQIIENSIDGSKLNHLIITPSCGSGQNIMTTF 1009  
QY 1021 TVIAVHYLDETEQWEKLEKQKQGALELIIKKGVTQQLAFQPSAFAAFKAPSTWLTA 1080



```

Db 1010 SVIATYLDATGQWENLGVDETRAIKQIMTGVAQWYKKAHSDHSAATNRASSSWLTA 1069
Qy 1081 YVVKVSLANVLLIA-IDQVLGAVKVMILLEKQKPGDGVFOEDAPVHQEMIGELNNNEK 1139
Db 1070 YVVKVLANASNNKXDSHEIICGGVKKWLIINROQPDGVFXENAPVJHGMELGGTK-GABP 1128
Qy 1140 DMALTAFLVLSIQAKDICEBQVNSLPGSITKAGDFLEANNMQLQSYTVAJAGVALAQM 1199
Db 1129 BASLTAFTVLTALLESRSVCKEQINILDOSSINKATDYLKKYKQLQPYTALTAYALAA 1188
Qy 1200 GELKPLINKLTTAKDKNRWDEKGLYNVEATSTALLALLQLQDFVPPVWRWLNQ 1259
Db 1189 DRLNDNRV-LMAASTGRNREWEYNATHNIEGTSVALLAKKKEFAEVGVFWRLIDQ 1246
Qy 1260 RYGGYGSTQATFMVFOALAQYQKADPDEQLNLDVSLQLPERSKITHRIHWESASLL 1319
Db 1247 KYGGYGTQATVMVFOALAAEYEQMPTHDNLNLSIKLPXEVPERYSNDRNAVQA 1306
Qy 1320 RSEETKENGEGTVTAEGKGQGLSVVTMYHAKA-DOLTCNKFDLVKTIKPAPEKEBPQ 1378
Db 1307 RVETKLNEDTVSASGGKATMTLTVYNAQREDANCKRPHLDVSVENVELNKKQAK 1366
Qy 1379 DAKNTMILEICTRYGQDQATMSILDISMNTGAPDPTDDLQKANGVDRIYSKIELDKAF 1438
Db 1367 GCKAALRLKICTRYLGEVDSTMTIIDISMLTGFPPDAEDLKLNSGVDRYISKFEIDNM 1426
Qy 1439 SDRNTLIYLDKVSHEDDCLAFKVOYENVELIQGAVKVAAYNLESCRFVHPPEKE 1498
Db 1427 AQKGTIVYLDKVSHEDECHFKIHXHFVGTIQGSKVYSYNLDSCQTKFYHPPEKE 1486
Qy 1499 DPKNLKCRDELRCABENCFIQSDDKVTLEBLDKACBPVGVYVYKTRLVKVLQSLNDF 1558
Db 1487 TGLLNKICHGNICRABETSLLNQKKIDQLRIQKACQNVYVYKTKLRISEKDN 1546
Qy 1559 DEYMAISQTIKSSDEQVQOQRTFSPIKCEALKLEKKHVMWGLSSDFWCKEKNL 1618
Db 1547 DIYFMDVLEVTKGTDNRNAQAKARQYVSQRKCEALNKLNDNLYLWGLSSDLNPKWDDI 1606
Qy 1619 SYIKGDTWVHNPDEECQDEENQKQCDLGAFTSMVVFQGP 1662
Db 1607 SYLITKNTWIERWNEDECCQDEBFQNLCDPFAQLSNTLTIFGCP 1650

RESULT 10
Q91132 PRELIMINARY; PRT; 1642 AA.
AC Q91132;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cobra venom factor precursor.
OS Naja kaouthia (Monocled cobra) (Naja siamensis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Naja.
OX NCBI_Taxid=8649;
RN (1)
RN SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RX MEDLINE=95108041; PubMed=7809120;
RA Fritzinger D.C., Bredehorst R., Vogel C.W.;
RT "Molecular cloning and derived primary structure of cobra venom
factor.";
RT factor.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:12775-12779(1994).
RN (2)
RN SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Fritzinger D.C.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; U09969; A268989.1; -.
DR HSSP; P01026; 1QQF.
CO GO:0005576; C:extracellular; IEA.

```

```

DR GO:0004866; F:endorpeptidase inhibitor activity; IEA.
DR GO:0006956; P:complement activation; IEA.
DR GO:0006954; P:inflammatory response; IEA.
DR InterPro; IPR002890; A2M N.
DR InterPro; IPR009048; AM_receptor_bind.
DR InterPro; IPR000020; Anaphylatoxin.
DR InterPro; IPR001840; Anaphylatoxin.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR008964; Invasin_intimin.
DR InterPro; IPR001599; MacroglobinA2.
DR InterPro; IPR001134; Netrin_C.
DR InterPro; IPR008930; Terp_cyc_toroid.
DR InterPro; IPR008993; TIMP_like.
DR Pfam; PF00207; A2M; 1.
DR Pfam; PF01835; A2M N; 1.
DR Pfam; PF01821; ANATO; 1.
DR Pfam; PF01759; NTR; 1.
DR PRINTS; PR00004; ANAPHYLATOXN.
DR ProDom; PD003264; Anaphylatoxin; 1.
DR SMART; SM00104; ANATO; 1.
DR SMART; SM00643; C34SC; 1.
DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR PROSITE; PS01178; ANAPHYLATOXIN 2; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00189; NTR; 1.
KW Signal.
FT SIGNAL 1 22 Potential.
FT CHAIN 1264 1642 beta chain of cobra venom factor.
FT CHAIN 23 648 alpha chain of cobra venom factor.
SQ SEQUENCE 1642 AA; 184516 MW; 2A71B2BD61D612A3 CRC64;

Query Match 49.1%; Score 4224; DB 2; Length 1642;
Best Local Similarity 49.6%; Pred. No. 8.2e-243;
Matches 824; Conservative 324; Mismatches 478; Indels 36; Gaps 20;

Qy 9 LALLLLTHPLALGSPMSIITPNTLRLESSETMVLAEHDAQGDVPVTVVHDPGKKLV 68
Db 8 LVAAALIGPGSGHALYTLIIPAVRTDTEQILVEAHGSDTPKQIDFVHDPKQKT 67
Qy 69 LSSEKTVLTPATNHMGNTFTIPANRPFKSEKGRNKFVTVQATGTQV-VEKVVVLSQS 127
Db 68 LFQTRVDMNPAGMLVTFETIEIPA-KEVSTDSRQNYVWQVT-GPQVRLEKVVLSYQS 125
Qy 128 GYLFIQTDKTIYTPGSTVLYRIFTVNHKLLPVGRVWVNIENPEGIPVKQDLSQQNLG 187
Db 126 SFLFIQTDKGIYTPGSPVLYRVFSMDHNTSKNKTIVIEFQTPEGILVSSNVD- 179
Qy 188 VLPLSW--DIPELVNGQWKIRAYYENSPOQVFSTEFVEKVEYVLPSEFVIEPEKPYI 245
Db 180 -LNFFWPYNLPDLVSLGTWRIIVAKYHSPEN-YTAYFDVRKYVLPSEFVRLQSEKPYI 237
Qy 246 YNEKGLVITITARFLYKKGVEGTAVFIFGIODGEORISLPESLKRIPIEDGSGWVLSRK 305
Db 238 DGNENFHVITARYLYGEEVGVAVFVGVKIDDAKKSIPDSLTRIFIIDGDKATLRD 297
Qy 306 VLLDGVNQPRADLVKKSILYVSATVILHSGSDMQAERSGIPVTSPYQIHFTKPYFK 365
Db 298 TFRS--RFPNLNELVGHLYASVTWVTESGDMVWVTEQSGIHIVASPYQIHFTKPYFK 355
Qy 366 PGMFPDLVMFVTPNPDGSPAYRVPVAVQGEDTVQSITQGDGVAKLSINTHPQKLSITVR 425
Db 356 PGVPYELTVYVTPNPDGSPAAHPVYVSEAFHSM-GTLLSDGTAKILINFLNAQSLPIVR 414
Qy 426 TKQQLSEAEQATRWQALPYSTVGNNSNNYLHSLVRLTELPGETLNVNLLRMDRAHEA 485
Db 415 TNHGDLPRERQATKSMATAYQTGGSGNYLFAVITSTEIRKPGDNLPNFNVKGNANSLK 474
Qy 486 KIRYTYLIMNKGRLLKAGRVREPGQDLVWLVPLSITDTDFPSFRLVAYVTLIGASQRE 545
Db 475 QIKYFYLLINKGKIFKVGQRPDRDQGNLVMTNLHITPDLIPSPFRFVAYYQV---GNNE 530
Qy 546 VVADSVWVDVKDSCVGLSVKSGQSDRQPVPGQOMTLKIEGDHARVWLVAVDKGVVL 605

```

Db 531 IVADSVWVDVKTCMGTLVVK---GDNLIQMPGAAMKIKLEGDPGARVGLVAVDKAVTVL 587  
Qy 606 NKNKLTOSKIWDVVEKADIGCTPGSGKDYAGVSDAGLTTSSSGQOQTAQAELOCPQ 665  
Db 588 NDKYISQAKIWTIEKSPGCTAGSGQNNLVFEDAGLAUTTSNLTKRSAAKCPQ 647  
Qy 666 AARRR-SVOLTEKRMKVGVY-PKELRKCCEGDNEMPMRFSQORTRFISLGEACKV 723  
Db 648 ANRRRSVLLDSNAKAAEFQDQRLKCCEDVHMENPMGYTCEKRAKYIQEODACKAA 707  
Qy 724 FLDCNVITLRRQHARASHGLARSNLEDIIAENIVSRSEPPESWLNVEDL-KEPP 782  
Db 708 FLECRVYKVRDNOQRESEFLARDNEDGFIADSIISRDFPKSWLNTKLTBEPN 767  
Qy 783 KNGISTKLMMNIFKDSITTWELLAVMSDKKICVADPFVTVQDFFIDRLRLYSVVRN 842  
Db 768 SQGISSTKMSFYLRDSITTWVLAVSFTTKGICVAEPYIRVMKVFFIDIQMPSYVKN 827  
Qy 843 EQVEIRAVLYNQONELKVRVELLNPAFCSLATTKRRHOOTITIPPKSSLSVYVIVP 902  
Db 828 EQVEIRALLNY-VNEDIYVRVELLNPAFCASASTGQRYQQPFIKALSSRAVFFIVP 886  
Qy 903 LKTLQVEVKAAYVHHFISDGVKSLKVVPEGIRNKTVAVRTLDPRLGRBQVQKEDI 962  
Db 887 LEQGLHDEVEIKASVQALMSDGVKRLKYVPEGVQ-KSIVTIVKLDPRKGVGTQLEVI 945  
Qy 963 PPADLSQVPTSETRILLQSTPVAQMTEDAVDAERLKHIVTPSGCGEQNMIGMPTV 1022  
Db 946 KARKLDRVPTETIKIILQDPAQIENSIDGSKLNHLIITPSGCGEQNMIRMAVP 1005  
Qy 1023 IAHYLDSTQWKEKLEKQGALELIKGYTQQLAFQPSAFAAFVRAVPSWLTAYV 1082  
Db 1006 IATYLDSTQWKEKLEKQGALELIKGYTQQLAFQPSAFAAFVRAVPSWLTAYV 1065  
Qy 1083 VKVPSLAVNLIA-IDSQVLCGAVKVLILEKQPDGVQFQEDAPVHQEMIGLRNNKNNK 1141  
Db 1066 VKVPSLAVNLIA-IDSQVLCGAVKVLILEKQPDGVQFQEDAPVHQEMIGLRNNKNNK 1124  
Qy 1142 ALTAFLVLSIQEAKDICEQVNSLPQITKAGDFLEANNMLORSYTVIAGYALAAQMR 1201  
Db 1125 YLTAFLVLSIQEAKDICEQVNSLPQITKAGDFLEANNMLORSYTVIAGYALAAQMR 1184  
Qy 1202 LKGPLNKLFTAKDKNRWEDPGKQVLYNVEATSYALLALLOLQKDFVFPVVRVNLQRY 1261  
Db 1185 LNDRLVMAASTGRD-HWEYVNAHTNIEGTSYALLALLOLQKDFVFPVVRVNLQRY 1242  
Qy 1262 YGGYGSQTATFVFOALAQYKQADPHQELNLDVSLQIPSRSSKITHRIHWSASLLRS 1321  
Db 1243 YGTYGQTATVAFQALAEYEQMPTHKDLNLDITIELPREVPIRYRINYNALLART 1302  
Qy 1322 EETKENGFTVTAEGKQGTLSVVTMYHAKAQDL-TCKNFDLKVITKPAPETEKRPQDA 1380  
Db 1303 VETKLNLDITVTAEGKQGTLSVVTMYHAKAQDL-TCKNFDLKVITKPAPETEKRPQDA 1359  
Qy 1381 KVTMLBICRYRGDODATMSILDISMTGEPADDDLLKOLANGVDRVYSKYVLDKAFSD 1440  
Db 1360 KGALMLKICRYLGEVDSTWIIIDISMLTGFIPDAEDITRLSKGVDRVYSKYVLDKAFSD 1419  
Qy 1441 RNTLIIVYLDKVSHEDECDLAFKQVQVFNVELIQGAVKYAYYNLEBSCFTHYFPEKEDG 1500  
Db 1420 KVAIVYLVNKSHEDECDLAFKQVQVFNVELIQGAVKYAYYNLEBSCFTHYFPEKEDG 1479  
Qy 1501 KLNKLCRDELRCACENCFQKSDDKVTELEERLXACEPGVDYVYKTRLVKQVLSNDFDE 1560  
Db 1480 LLNKICIGNVCRGAGETCSLNNHQBIDVPLQIEKACETNVYVYKTRLVKQVLSNDFDE 1539  
Qy 1561 YMAIEGTIISGSDEQVQOQRTFISPIKCRALXLEBKHYLMWGLSSDFWQKPNLSY 1620  
Db 1540 YVMDVLEVIQGTDENFRKTHOYISQKQCBALNKNVDDYLVNGSRDILLPTKDISY 1599  
Qy 1621 IIGKDTVWEHWPBEDECOBENKQCOQDLGATESNMVFGCP 1662  
Db 1600 IITNTWIERWHPBEDECOBENKQCOQDLGATESNMVFGCP 1641

RESULT 11  
Q91588 PRELIMINARY; PRT; 1589 AA.  
ID Q91588;  
AC Q91588;  
DT 01-NOV-1996 (TREMBlurel. 01, Created)  
DT 01-NOV-1996 (TREMBlurel. 01, Last sequence update)  
DT 01-MAR-2004 (TREMBlurel. 26, Last annotation update)  
DE Complement component C3 (Fragment).  
CS Xenopus laevis/gilli.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_taxid=8359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Lambiris J.D., Pappas J., Mavroidis M., Wang Y., Manzone H.,  
Schwager J., Du Pasquier L., Silibovsky R.,  
RT "The third component of xenopus complement: cDNA cloning, structural  
and functional analysis, and evidence for an alternate C3  
transcript.";  
RT Eur. J. Immunol. 25:572-578 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Lambiris J.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U19253; AB060608.1; -;  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0004866; P:endopeptidase inhibitor activity; IEA.  
DR GO; GO:0006956; P:complement activation; IEA.  
DR GO; GO:0006954; P:inflammatory response; IEA.  
DR InterPro; IPR002890; A2M N.  
DR InterPro; IPR009048; AM receptor bind.  
DR InterPro; IPR000020; Anaphylatoxin.  
DR InterPro; IPR001840; Anaphylatoxin.  
DR InterPro; IPR008964; Invasin incrimin.  
DR InterPro; IPR001599; Macroglobulin.  
DR InterPro; IPR001134; Netrin C.  
DR InterPro; IPR008930; Terp cyc toroid.  
DR InterPro; IPR008993; TIMP like.  
DR Pfam; PF00207; A2M; 1.  
DR Pfam; PF01835; A2M N; 1.  
DR Pfam; PF01821; ANATO; 1.  
DR Pfam; PF01755; NTR; 1.  
DR PRINTS; PR00004; ANAPHYLATOXN.  
DR ProDom; PD003264; Anaphylatoxin; 1.  
DR SMART; SM00104; ANATO; 1.  
DR SMART; SM00643; C345C; 1.  
DR PROSITE; PS00477; ALPHA 2 MACROGLOBULIN; 1.  
DR PROSITE; PS01177; ANAPHYLATOXIN 1; 1.  
DR PROSITE; PS01178; ANAPHYLATOXIN 2; 1.  
DR PROSITE; PS50189; NTR; 1.  
FT NON\_TER 1.  
FT CHAIN 592 1489 complement component C3 alpha chain.  
FT CHAIN <1 588 complement component C3 beta chain.  
SQ SEQUENCE 1589 AA; 177904 MW; DCB777FB4B11456A CRC64;  
Query Match 48.9%; Score 4206.5; DB 2; Length 1589;  
Best Local Similarity 51.3%; Pred. No. 8.6e-242;  
Matches 817; Conservative 288; Mismatches 465; Indels 23; Gaps 18;  
Qy 83 MGVTFPIPNRFEKSEKGNKEVTCATFGCVTKVLSQSGYLFIQDTKIYTPG 142  
Db 8 LGTAVKIPSGDSKDTKSK-QVYVYVSKNVCALEKVVLSYHSGYIFLQDTKIYTPG 66  
Qy 143 STVLYRIFTVNHKLLPVGRVTWVNIENPEGIPVKQDSLSQNLGVFLSWDIPELVNMG 202  
Db 67 STVLYRIYVMYKQPIKTLIEFTUPDGIIVNTHLQDSKGIISQSHKLPALLSLG 126

QY 203 QMKIRAYVENSPOQVSEPERVEKEVYVLPSEVIEVTEPEKEXYLYNEKGLVETITARFLYG 262  
Db 127 VWTISAKEDTPOQNTTINFKEVYVLPSEIILKPDNTNIFY-SRASEFGVLIQOYLXG 185  
QY 263 KKVETGATVIFIGIDGEGORISLPESLKRPIPIEDSGEVLVSRKVLIDGQVONPRAEDLVGK 322  
Db 186 KHVDDYAFVLFGLKMDNVKIGIPESLARIIDDBEGRAELKRLDLVKYFEKP--EDMLQF 243  
QY 323 SLVVSATVILSHSGDMVAERSGPIVTPSPQIHFTKTPKYPKPGPPFOLMVFVTPNPGS 382  
Db 244 SLVTVSVVVTGSDMVAEAEIHSIVTTPYKILFKTKSKYFKPGPPFDMVYVTPNPGS 303  
QY 383 PAYRVPVAVQESDVTQSDGVAKLSINTHSQKPLSITVTKKQELSEABQATRTWQ 442  
Db 304 PARRIRVVVE-FGNVEGTOGDGTRULTANPANINLLPITVTRDTLPPGCRQATNT 362  
QY 443 ALPYSTVGNSSNNYHLVRLTEAPGETLNVNLLR-MDRAHEAKIRYVTVYILMKNKGRLL 501  
Db 363 ATAYRPSRAQGNLYLHISAGSDIRPGENIPVNFIRNTDAGVQKQIEHFTYLLISGRIL 422  
QY 502 KAGROVREPQDLVVLPLSITTDTPISFRVAYYTLIGASQREVVDVSDVVDXDSVCG 561  
Db 423 KVGROQROGQPFVFTMSIPVDTLIPSFRIVAYY-IVSAGGGRDIVSDSLVDDVDDCMG 481  
QY 562 SLVKSQSGSEDRQ-PVPGQOMTLKIEGDHGARVVLVADVGVLNKKKQLTQSKIWDVY 620  
Db 482 TLSVTGDKDRDNTIQTGSPNRLKLRADHKAYVGLVAVDKGVVYVLSKFKITOKKWDGV 541  
QY 621 EKADTGTCPGSKDVAGVPSDAGITFTSSSQOQTAORAELOCPQPARBR--SVOLTEKR 679  
Db 542 EKSDIGCTPGSGANSEGVSDAGLALQTSFGINTAQNSDAQCPAPARRRSGSVATEIK 601  
QY 680 MDVKGYKPELKKCCEDGMRENPMRFQCRTRFISIGEACKVFLDCCNYYATELRQHA 739  
Db 602 SGKSEYKDKAKKCCDGMQENLGMCHTCDRRARYLDGECVDAFVDCCKY-EKKREAE 660  
QY 740 RASHLG--LARSNLDIEDIIASENTVSSSEPEKSNVVEDLKEPPK-NGISKLNNIPLK 796  
Db 661 KLSKDDTLGSDDEDETLNDSDIVSRTEPSEFWKVBQMKPDVNGISKLTLNVFK 720  
QY 797 DSITWEILAVMSDKKIGCVADPFVTVMDFFIDRLPYSVVRNEQVEIRAVLVNYRQ 856  
Db 721 DSITWEILAVLSNKLGVQGYEIKWMDFFIDKLFYSVVRNEQVEIRAILVNYR- 779  
QY 857 NOELKRVRELLHNPAFCSLATTERRHOQTTIPPKSSLSVPYVIVPLKTGLQREVEKAAV 916  
Db 780 NDRIKRVRELLHNPEFCSLSTAKKXQEVWIGALSSIAVPVIVPLTGLQHDIVKASV 839  
QY 917 YHH---FISDGVKSLKVVPEGIRKNTVAVRTLDPERLREGVQKEDIPPDLSQVDP 973  
Db 840 SAQSGFFGADGVYRKKLVVPEGMRIACQVKTIVLEPEVKGKGVQSEKVEALNPKNVPR 899  
QY 974 TESETRILQGTPEVAQMTEDAVDAERLKLHIVTPSGGCEQNMIGMTPTVIAHYLDETEQ 1033  
Db 900 TIDITITLQGTPIQWVEDAIDGNMNNHLLIVPAGGCEQNMISTPSVIATRYLDASQ 959  
QY 1034 WEKFGLEKROGALEILKGYTOQLAPQPSAFAPFKRAPSTWLTAYVVKVPSLAVNLI 1093  
Db 960 WERVGNRDRQALKNMROGYAQAFKPDNSAAWDXDRPASTWLTGYVAKVFGNAQEF 1019  
QY 1094 AIDSOLGCAVKWILILEKQKDPGVQEDAPVTHQEMIGGLRNN-NEKMALAFVLISIQ 1152  
Db 1020 DIEANVLCSSVKWILILEKQKDPGLFQENAPVTHQEMVGGITGAEGDSSLTAFILIAML 1079  
QY 1153 EAKDICEBVNSLPSITKAGDFLEANYMNLORSYVTAIAGYALACQGRILKGLINKFLT 1212  
Db 1080 EQCRTCNEHNNLQVSDIKASSYLVGGYPLKPKYSITATSALALAGKL--PNTNKLUS 1137  
QY 1213 TAKDKNRWEDPGKQLYNVEATSALLALLQKDPDFVPPVVRWLNQRYVGGYGTQAT 1272  
Db 1138 ASIGNTHWEEPKRIFISLETTSYALLTLKMKEDFLTGGIVRWLNQRYVGVYGTQAT 1197  
QY 1273 FVWFQALAQYQKDAEDHDELNDVSLQLPSSRSKITHRIWESASLLRSEETKENGFTV 1332

Db 1198 IVMFQALAQYQDIDISVNMELNDVSLHLPERQOPLTYVINHENALLARSATRLNQDFV 1257  
QY 1333 TAEGKQOGLSVVTVYHAKAKD-QLTCKNFKDLKVTIKPAPETERPODAKNTMLEICTR 1391  
Db 1258 KAKGKQOGLSVVTVYHAKAKD-QLTCKNFKDLKVTIKPAPETERPODAKNTMLEICTR 1316  
QY 1392 YRDODATMSIILDISNMTGFAEDTDLLKQLANGVDRYISKVELDKAFSDRNTLIYLDKV 1451  
Db 1317 HLKNDVATMSIILDISNMTGFAEDTDLLKQLANGVDRYISKVELDKAFSDRNTLIYLDKV 1376  
QY 1452 SHSEDDCLAFKVOHVENVELTOPGAVKYVYNNLEESCTREYHPEKEDGKLNKCRDEL 1511  
Db 1377 SHLEECRFYAHQFPEVGFQIPASVTYDITTDNRCTKSHVEESALGRICQGDLC 1436  
QY 1512 RCAEENCFIQKS-DOKVTLEERLDCACBPQVYVYKTLVQVLSNDFDEYIMAEOTIK 1570  
Db 1437 RCAEENCFIQKS-DOKVTLEERLDCACBPQVYVYKTLVQVLSNDFDEYIMAEOTIK 1496  
QY 1571 SGSDVQVQOQTFISPIKREALKLEKKHYLWGLSSDFWGEKPNLSYIIGKDTWVEH 1630  
Db 1497 QGDEDEPKTNFISHIKCRKALNNQNRDYLWGTGDLWRQPDGYSYIIGKDTWVEH 1556  
QY 1631 WPEDECOBENKQOQDILGAFETSMVYVFGCPN 1663  
Db 1557 WPNRECOBENKQOQDILGAFETSMVYVFGCPN 1589

RESULT 12  
QYIYA9 PRELIMINARY; PRT; 1642 AA.  
ID QYIYA9 PRELIMINARY; PRT; 1642 AA.  
AC QYIYA9  
DT 01-MAY-1999 (TREMELrel. 10, Created)  
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)  
DB Complement C3-H1.  
OS Cyprinus carpio (Common carp).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Cyprinus.  
OC NCBI\_TaxID=7962;  
RN [1]  
SEQUENCE FROM N.A.  
RC TISSUE=Hepatopancreas;  
RX MEDLINE=20203443; PubMed=10741402;  
RA Nakao M., Mutsuro J., Obo R., Fujiki K., Nonaka M., Yano T.;  
RT "Molecular cloning and protein analysis of divergent forms of the  
RT complement component C3 from a bony fish, the common carp (Cyprinus  
XL carpio): presence of variants lacking the catalytic histidine.";  
DR Eur. J. Immunol. 30:858-866 (2000).  
DR EMBL; AB016211; BAA36619.1; -.  
DR HSP; P01026; IQQF.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0004866; F:endorpeptidase inhibitor activity; IEA.  
DR GO; GO:0006956; P:complement activation; IEA.  
DR GO; GO:0006954; P:inflammatory response; IEA.  
DR InterPro; IPR002890; A2M\_N.  
DR InterPro; IPR005048; A2M\_receptor\_bind.  
DR InterPro; IPR000020; Anaphylatoxin.  
DR InterPro; IPR001840; Anaphylatoxin.  
DR InterPro; IPR001599; MacroglobulinA2.  
DR InterPro; IPR001134; Netrin\_C.  
DR InterPro; IPR008930; Terc\_cyc\_toroid.  
DR InterPro; IPR008993; TIMP\_like.  
DR Pfam; PF00207; A2M; 1.  
DR Pfam; PF01835; A2M\_N; 1.  
DR Pfam; PF01821; ANATO; 1.  
DR Pfam; PF01759; NTR; 1.  
DR PRINTS; PR00004; ANAPHYLATOXN.  
DR ProDom; PD003264; Anaphylatoxin; 1.  
DR SMART; SM00104; ANATO; 1.  
DR SMART; SM00643; C345C; 1.  
DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.





AC QSVIAT;  
 DT 01-MAY-1999 (T-EMBLrel. 10, Created)  
 DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
 DB Complement C3-S.  
 OS Cyprinus carpio (Common carp).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Cyprinus.  
 OX NCBI\_TaxID=7962;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hepatopancreas;  
 RX MEDLINE=20203443; PubMed=10741402;  
 RA Nakao M., Mutsuro J., Obo R., Fujiki K., Nonaka M., Yano T.;  
 RT "Molecular cloning and protein analysis of divergent forms of the  
 RT complement component C3 from a bony fish, the common carp (Cyprinus  
 RT carpio): presence of variants lacking the catalytic histidine";  
 RL Eur. J. Immunol. 30:858-866(2000).  
 DR EMBL; AB016213; BAA36621.1; -;  
 DR HSSP; P01026; 1QQF.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.  
 DR InterPro; IPR002890; A2M.N.  
 DR InterPro; IPR009048; AM.Receptor bind.  
 DR InterPro; IPR000020; Anaphylatoxin.  
 DR InterPro; IPR008964; Invasin.  
 DR InterPro; IPR001599; Macroglobulin.  
 DR InterPro; IPR001134; Netrin.C.  
 DR InterPro; IPR008930; Tercp\_cyc\_coroid.  
 DR InterPro; IPR008993; TIMP\_like.  
 DR Pfam; PF00207; A2M; 1.  
 DR Pfam; PF01835; A2M.N; 1.  
 DR Pfam; PF01821; ANATO; 1.  
 DR Pfam; PF01759; NTR; 1.  
 DR ProDom; PD003264; Anaphylatoxin; 1.  
 DR SMART; SM00104; ANATO; 1.  
 DR SMART; SM00543; C34SC; 1.  
 DR PROSITE; PS00477; ALPHA 2 MACROGLOBULIN; 1.  
 DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.  
 DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.  
 DR PROSITE; PS01189; NTR; 1.  
 SQ SEQUENCE 1646 AA; 184194 MW; BB40BE1DE08A9390 CRC64;  
 Query Match 40.9%; Score 3521; DB 2; Length 1646;  
 Best Local Similarity 44.1%; Pred. No. 8e-201;  
 Matches 734; Conservative 323; Mismatches 569; Indels 38; Gaps 24;  
 9 LLLLLLTHPLALGSPWYIIIPNLRLESEETVLEAHDAGQ-DVPVTVTVHDFPGKLL 67  
 8 LTVVLSSPLLTLCDPLFVLSAPNLRVGSSENVFEAQDYSGRAIEVKIIVKHPKDR 67  
 68 VLSEKTVLTPANMGVNTFIPANREPKSEKGRNK-FVTVQATFGTVQVKEVVLVSLQ 126  
 68 EILSQSVRLTADNFQIIRKIPDDQNYFSPNPLEKQYVYLOAQFPSTLEKEVMSFQ 127  
 127 SGVLFQTDKTIITPGSTVLYRIFTVNHKLLPWGR-T-VAMNENPEGIPVAKQDSLSONQ 185  
 128 SGVIFVQTDKPIITPASTVOYRIFSLTPNLEPLSQSGITVEIMNPQGITVSEKIFPVK- 186  
 186 LGVLPLSWIDPELVNNGWKIRAYENSQQVSTFEFVEKVELVLPSEFVIEPEKEFYI 245  
 187 -GMKSGKYPFEIASPGIMKVVTLFNSNTFQKFTADFEVKEVLPSTFEVKLKPSSFFV 245  
 246 YNEKGLVETITARFLYKGVKVEGTAFVIFIGIQGEORISLPSLRKRIPIEDSGSGWLSRK 305  
 246 -GDSPLTVDLIEAKYLGQVDGNAFVVGVMNGEKKSIFPSLQVQVIRREGTAELTSG 304  
 306 VLLDGVNPRADLVGKSLVYNSATVILHSGSDMVAERSGPIVTSFYCHFTTKPKYFK 365  
 305 MITKTF--PNINLVGQSIYVSVLTSGESSEKVAERRGQIVTSPYSIHFKTKTOFFK 362  
 366 PGMFFDLWVFTNFDGSPAYRVPVAVQG-EDTVQSLQDGDGVAKLSINTHPSQKPLSITV 424

Db 363 PGMFPDYSVYVTPDQTPAVNVEVVGSGQTVKGOTKINGIAKVTVNTLGSSSTQELITA 422  
 QY 425 RTKQELSEBOATRTMOALPY-STVGNNSNYHLVSLVRLTELEPGETLVNVLAMDRAH 483  
 Db 423 KTKQFQLRDQQAQVKKVTAHAYTPKASKKNYVHIGIDAAELQMGDSMKV--FLMTGOSP 480  
 QY 484 EAKRYVYLYIMNKGRLKAGRVQRPQDLVVLPLSITTTDFIPSPRLVAYTYLIGASQ 543  
 Db 481 GVKQDITYMILSKGQIVSDVRFKR-GQSLVSLVPVTKMVPSFVAYVHV---GS 535  
 QY 544 REVVADSVVDVKDSVUGSLVWKSQSESDROPV--PQQVTLKIEGDHARVVLVAVDKG 601  
 Db 536 SEVVSDSVVDVKDTCNGKLQVK---VKDQNTYDTGDGEVKLEITGDPKARKVGLVVDKA 592  
 QY 602 VFVLNKNKLTQSKIMVDVEKADIGCTPGSGKDYGAVFSDAGITFTSSSGQOQTAQRAELQ 661  
 Db 593 VOVLN-KNRLTQTQIMDVIEKHDTGCTAGGRDRMGVFTDAGLWFSNTAGTNTIRTP 651  
 QY 662 CPQPAARRRSVQLTERKMDKVKYKPKELKCCEDGRENPMRPFSCQRTTRISIGEACK 721  
 Db 652 CPKPKRRRAESLLKITSTLAEKYGELKQCCVGMRENKLGITCDRRSRVADGKECV 711  
 QY 722 KVFDDCCNYITELRRQCHARASHGLARSNLDDEIIAE-ENIVSRSEFPESWLVNVEDLKE 780  
 Db 712 DAFLHCCNQMKTKDKVDEVMVLAESDDDDDYTESSEIVSRITQFPESMLWEEIDLCD 771  
 QY 781 PPKNGISTKLNNIPLKDSITTWILAVSMDSKXGICVADPPFVTVMQDFIDRLRPSYV 840  
 Db 772 --KCAIPTKEKAIYVKDSITTWQLAISLPTTGICVABPEEIVFKSLFIDLKKMPYSAV 829  
 QY 841 RNEQVEIRAVLYNRQNELKVRVELLHNPFAFGLSLATTKRRHQITITIPKSSLSVPYI 900  
 Db 830 RGEQLERAILHNVTPKKQ-KRVVEFMETEDVCSSASKGKYRTTVSDVKSSISVSIVI 888  
 QY 901 VPLKTGLQVEYKAAVTHFISDGVKSLKVPPEGIRMNKTAVAVRTLDERLGRGV-QK 959  
 Db 889 IPMTLGNMIEVKASAYDSYTDGVRKPLKVAEGVLVPLQRLQNELNPAKNGEPILK 948  
 QY 960 EDIPADLSQDVPDTESETRILLQGTFFVAQMTDAVDAERLKHILVTPSGCGEONMIGT 1019  
 Db 949 GEIP---ADRPDTPANTVISITGEEISQVBOAISGDFMGKLLVQPSGCGEONMIGT 1004  
 QY 1020 PTVAHVHLDTEQWEXFLEKQGALELILKGYTQQLAFROPSSAFARVBAESTWLT 1079  
 Db 1005 LPVATHYLDNTQWEANGIQRTTEALNHIRTGYQLTKRSDGSYAAFTSRPSTWLT 1064  
 QY 1080 AYVVKVSLAVNLIAIDSVLQCGAVKWLILEKQKPDGVPQEDAPVTHQEMIGLNNNEK 1139  
 Db 1065 AYVAVKPAIANNLVTIEENVICSAKWLVLHKQLPDGSKEDSAVSSGVMGVDVQGN-A 1123  
 QY 1140 DMALTAFLVLSLOAKDICEQVNSLPGSITKAGDFEANYMKNLQSRYSVTAAGVALAQ 1199  
 Db 1124 DASLTAFLVIAQGEIEICAGSVASQESIRKAVSLEGRLPQLTNPYAVAMTSVAMASE 1183  
 QY 1200 GRKGLPLNFKLTAKDKRWEDFGKQLYNVEATSALLALLQLKDFDFVPPVFWRLNEQ 1259  
 Db 1184 NKLKNDILMRHSTQOEASRSMTVPVGHHSLEATAVALVRDRDFDAGEAVHWLARQ 1243  
 QY 1260 RYVGGVGSCTAATFWFCALAQYKADPHQELNLDVLSQLPSRSSKITHRIHWSASLL 1319  
 Db 1244 QAHYGYGTTQATLMVFQVAEYNTQVKTQNFNLDELVSAGRSKPVWTKRDNALH 1303  
 QY 1320 RSEETKENEGTVTAEGKGQGTLSVVTMHAK-AKDQLTCNPKDLKVTIKPAPETEKRPQ 1378  
 Db 1304 RSDRVEINKDPSVTARGTKATLSVLTLYYAREAVKSKDCTLLFDL--TVKMEEDNKSNO 1361  
 QY 1379 DAKNTMILEICTRYRGD-QDATMSILDISMTCTAFAPDTDLKQLANGVDRYT-SKVELDKA 1437  
 Db 1362 GAITYKFTMPFYKSKKKAATMTLLDGIPTGFSVDSRDLQELSTGKERYIQKEMDKV 1421  
 QY 1438 FSDRNTLIYLDKVSHESDCLAPKHQYFNVELIPGAVKVVAYVYNNLSECTRFVHPBK 1497

Db 1422 LSERGLILYLEKVLHREKERIAPRMHKKIQNVGLLOPAAVTIYEVYSPDARCTKPPHPR 1481

Qy 1498 EDGLKLNKDELRCRAENCFOKSDDKYTLBERLDKACEPGVDVYKTRLVKVLQSLND 1557

Db 1482 EDGLYLRUCKDMCCAEKNCYOKK-NHVEDDERFNKACEAGMDVYKVVGMGLKKD 1540

Qy 1558 FDEYMAIEQTKSGSDEVQOQRTTISPICKREALKLEEKHYLMWGLSDFWGEKEN 1617

Db 1541 SDIEMKVEQVLKSGTDEBEGVTFTLAPRSCREHLGLVKGSYLIMGKSVDLPLKGS 1600

Qy 1618 LSYIIGKDTWVHWPEDDECOENOKQCDLGAFTESVVFCC 1661

Db 1601 LOYINGEQTWVEYPTREESQTRHRRDYIGISDLENSURKEC 1644

RESULT 15

ID Q9YIA8 PRELIMINARY; PRT; 1640 AA.

AC Q9YIA8;

DT 01-MAY-1999 (TRENBLrel. 10, Created)

DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)

DE Complement C3-H2.

OS Cyprinus carpio (Common carp).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Cyprinus.

OX NCBI\_TaxID=7962;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Hepatopancreas;

RX MEDLINE=2020343; PubMed=10741402;

RA Nakao M., Mutsuro J., Obo R., Fujiki K., Nonaka M., Yano T.;

RT "Molecular cloning and protein analysis of divergent forms of the

RT complement component C3 from a bony fish, the common carp (Cyprinus

RT carpio): presence of variants lacking the catalytic histidine.";

EL Eur. J. Immunol. 30:858-866(2000).

DR EMBL; AB016212; BAA36620.1; -.

DR HSSP; P01026; 1QQF.

DR GO; GO:0005576; C:extracellular; IEA.

DR GO; GO:0004866; F:endorpeptidase inhibitor activity; IEA.

DR GO; GO:0006956; P:complement activation; IEA.

DR GO; GO:0006954; P:inflammatory response; IEA.

DR InterPro; IPRO02890; A2M\_N.

DR InterPro; IPRO03048; AM\_Receptor\_bind.

DR InterPro; IPRO00020; Anaphylatoxin.

DR InterPro; IPRO01840; Anaphylatoxin.

DR InterPro; IPRO01599; Macrogloblna2.

DR InterPro; IPRO01134; Netrin.C.

DR InterPro; IPRO08930; Terp\_cyc\_toroid.

DR InterPro; IPRO08993; TIMP\_like.

DR Pfam; PF00207; A2M; 1.

DR Pfam; PF01835; A2M\_N; 1.

DR Pfam; PF01821; ANATO; 1.

DR Pfam; PF01759; NTR; 1.

DR PRINTS; PR00004; ANAPHYLATOXN.

DR ProDom; PD003264; Anaphylatoxin; 1.

DR SMART; SM00104; ANATO; 1.

DR SMART; SM00643; C345C; 1.

DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.

DR PROSITE; PS01177; ANAPHYLATOXIN\_1; UNKNOWN\_1.

DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.

DR PROSITE; PS0189; NTR; 1.

SQ SEQUENCE 1640 AA; 183695 MW; 495B0845275BF159 CRC64;

Query Match 40.5%; Score 3489; DB 2; Length 1640;

Best Local Similarity 43.7%; Pred. No. 6.5e-199;

Matches 726; Conservative 327; Mismatches 568; Indels 42; Gaps 22;

Qy 9 LLLLLLTHLPALGSPHYSITPNIWLRSEETNVLEAHD-AQGDPVTVTVHDPGKKL 67

Db 8 LTVVLLSPLLTCDPLVLSAPNLRLRVGSSNMFVDAQYREDLNKXIIIVNHPKXDM 67

Qy 68 VLSEKTVLTPATNHMGNVTFTIPANREPKSEKGRNK-FVTVOATFGTQVVEKVVLSLQ 126

Db 68 DILSKSVLTAANCFOILTDIKIPDDQNFSDPLEKQVYLQAFPPSVTLKVVLSFPQ 127

Qy 127 SGYLFQTDKTIYTPGSTVLYIFRFTVNHKLLPVGR-TVMVNIENPEGIPVKODSUSSNQ 185

Db 128 SGYIFQTDKTIYTPGSTVLYIFRFTVNHKLLPVGR-TVMVNIENPEGIPVKODSUSSNQ 186

Qy 186 LGVLPISWDIPELVNMGOKIRAYENSPOQVSTFEFEVKEVLPSPFVIEPTKFFYI 245

Db 187 -GKSGNYALPEVASPGIWKLVTLFSNTQKFTADFVKEVYVLTFFVKLPSPSFYV 245

Qy 246 YNKGLEVTITAPFLYKVKVEGTAFVIFGIQGEORISLPESLRIPKIPEDSGEVLGRK 305

Db 246 -REQSLTVDIKAYLFGQNVDTAFVYFGWENKTRIPASLOKKKMKGEATLNTQ 304

Qy 306 VLLDGVQNPRAEDLVGKSLVVSATVILHSGSDMWQASRGIPVITSPQIHFITKPKYFK 365

Db 305 MITKTP--PNINQLVGRSIVVSULLTESGESEWAEERRGIQVITSPYTHFRKTPHFK 362

Qy 366 PGMPFDLMVFTNPDGSPAYRVPVAVOGEDTVQSITQDGVAKLSINTHPQKPLSIVR 425

Db 363 PGMPFVSVVYTNPDQTPAVRMEVEV-NPGGVKGRTRDNGIAKVMVNTPGSSTLEITAK 421

Qy 426 TKQSELSEABQARTMQALPYSTVGNMNYHLVLRTELRCETLNVNFLRMDRAHEA 485

Db 422 TKDPKLGDQQAQVTTWTAQAYIPKOGSNYHLIGIDAAELQIGDSMTVN--LNTQIPIGV 479

Qy 486 KIRYYTILMNKRLKAGRQVREPGQDLVPLSITTDFTPSFLVAYYILIGASQORE 545

Db 480 TQDQFTYMLSKGQIVKAVRFRQ-QQSLVTLVSFVTKDMPSFRFAYYHV---GSSE 534

Qy 546 VVADSVVMDVKDCVGLSVVVGQSDRQPV-PGQOMTLKIEGDHCAVVLVAVDKGVFV 604

Db 535 VVSDSIWVDKDCMGTLOIEVKHKEKINTYGNKVELQITGDFGAKVGLVVDKAVQV 594

Qy 605 LNKKNKLTQSKIWVVEKADICTPGSGKDYAGVTSADCLTPTSSSGQQTQARABLCQPQ 664

Db 595 LN-KNRLTQTKIWDVIEKHDTGTCTAGGSDSGVFTDAGLMEVSNAGTNTTRVPECFI 653

Qy 665 PAARRRSVQLTEKRMVKVYKPKELKCCEDGMENPMRFSQRRRTFISLGEACKVVF 724

Db 654 HAKKRAENLQITGLAGKSGELKQCCVGDGMDNKLGYTCERRAMVLDGECVRAF 713

Qy 725 LDCNVITELRROHARASHLGLARSNLD--IAENIVSRSEFPESLWNVDDLKEP 781

Db 714 LDCNCQMKRKNTETEEEMILARSDDDDDDDDYTDSEITSRTPESLWNEEVDLPTS 773

Qy 782 PKNGISTKLMNIFLKDSITTWELAVSMDSKKGICVADPFVTVMDQFFIDLRPLYSVVR 841

Db 774 DKGETTSIPKVIYKDSITTWQILAVSLSPITLIGICVAFPEEMAVFKQLFIDLKMPYSVR 833

Qy 842 NEQVEIRAVLYNTRQNELKVRVELLHNPAFGLATTYRRHQQTITIPKSSLSVYVIV 901

Db 834 GEOLFKAIHNYTPNKQ-KVRVELMKTEDVCSASKGKYRTVTVSDSGSISYSYVII 892

Qy 902 PLKTGLQVEVEKAAVYHHFISDGVRSKLVKVPGEIRMNKTVAVRTLDERLREGVQKED 961

Db 893 PMLTGNHDIKVRKASADVITYDGVKALKVSEGV-----LTSVHRENVELNPKRPF 945

Qy 962 IPADLSD-QVDETESEFILLQGTTPVAQMTEDAVDAERLKHILVTPSGCGEQNMIGWTP 1020

Db 946 IFKADIQDNRPDTPANTYISITGEEITQTVEQALSGSFMGRILVQPSGCGEQNMIGMTL 1005

Qy 1021 TVIAVHVLDETEQWEKFGLEKQGALELTKKGYTQQLAFROPSSAFAAFVRAFPSTWLTA 1080

Db 1006 PVIATHYLDSTQWETVGFERRNEAINHRTGYQLGVKRSKSGSYAAWIHRPSTWLTA 1065

Qy 1081 YVVKVPSLVANLIADSOVLCCAVKWLILEKQKPGVFOEDAPVHQEMIGGLRNNEKO 1140

Db 1066 YVAKVFAAMANNLTIEKVLCSALEWLVHRTQDGSFKESAVIHGEMVDVQ-GNDAD 1124

Qy 1141 MALTAFLVLSIQEAKDICEQVNSIPGSIITKAGDGFLEANVMNLQRSYTVTAIAGVALAQMG 1200



Db 1125 ASLTAFAVIAOFAEIBICAKE--SLHERIRKAVAVLEGRLSOLTNPYAVALTSYANANEN 1182  
Qy 1201 RLKGPLLNKFLTTAKDNWEDPGKOLYVNEATSVALALLQLKDFDFVPPVVRWNEOR 1260  
Db 1183 KLNKDILMKHSQRGAGRSWTVPQGHHSLEATAYAVLALYKAKDFKAGEAVHNLNRQ 1242  
Qy 1261 YCGGYGSTQATFMPQALAOYQKADAPDHQELNDVSLQSPRSKITHRIHWESALLR 1320  
Db 1243 SHYGGYGTQSTIMVFAVAYRTQMKQCNFNLEVELSVEGRSPRVYTIKKANAHLTR 1302  
Qy 1321 SEETKENEGFTVTAEGKGOTLSVVTWYHAKADOLT-CNKFPDLKVTIKPAPETEKPOD 1379  
Db 1303 SDRVDINQENVTARGTATLSVLTLYARPVEKSDCTFNLTV-----KMEKYNKG 1356  
Qy 1380 AKNTMILEICTRYRGO-DATMSILDI SMITGFAPD TDDLKOLANGVDRIYSKYELDKAF 1438  
Db 1357 AVASYKLTWDFIYKSDKTATMTILDIGLPTGFEVESDLKELSGSKERYIQKYENNKVL 1416  
Qy 1439 SDSNTLIIYLDKVSHSEDDCLAFKVHOFNVVELIOPGAVKVYAYYNLEESCTREYHPEKE 1498  
Db 1417 SERGSLILYLDKVSHKEKEVISFRMHQMLDVGLLOPAAVTIYEYSPDARCKEYHPERT 1476  
Qy 1499 DGKLNKCRDELICRAENCFTQKSDDKVTLLEERLDKACBPGVDYVYKTRLVKVQLSNDP 1558  
Db 1477 DGAIYRLCKGDLCYCAEENCYQKK-NRVSDEERENRACEAGMDYVYKTVVGMDLKQDS 1535  
Qy 1559 DEVYMAEOTIKSGSDEVQVGQORTFISPIKCREALKLEEKHYLMWGLSSDPWGEKPNL 1618  
Db 1536 DIYDMKVEQVLKEGTDVVEEKVRPFLARPSCREYLGLEKSYLNGRSVDLPRLRGS 1595  
Qy 1619 SYIIGKDTWVWHPEDEQDEENKQCCQDLGAFTESMVVFGC 1661  
Db 1596 QYVFGQTWVYWPTEESSTREHRYNGIAKLNLSLLSYGC 1638

Search completed: December 22, 2004, 00:37:16  
Job time : 253 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2004, 00:16:53 ; Search time 197 Seconds  
(without alignments)

3028.258 Million cell updates/sec

Title: US-09-875-519A-22

Perfect score: 8609

Sequence: 1 MGTSFGSLLLLLLHPLA.....KQCQLGAFTEMVVFGCPN 1663

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq\_23Sep04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	8609	100.0	1663	2	AAW34606	AAW34606 Wild type
2	8605	100.0	1663	2	AAW34607	AAW34607 Human C3
3	8605	100.0	1663	2	AAW34610	AAW34610 Human C3
4	8605	100.0	1663	2	AAW34609	AAW34609 Human C3
5	8605	100.0	1663	2	AAW34616	AAW34616 Human C3
6	8604	99.9	1663	2	AAW34618	AAW34618 Human C3
7	8603	99.9	1663	2	AAW40989	AAW40989 Human C3
8	8601	99.9	1663	2	AAW34608	AAW34608 Human C3
9	8600	99.9	1663	2	AAW34028	AAW34028 Human C3
10	8599	99.9	1663	2	AAW34627	AAW34627 Human C3
11	8598	99.9	1663	7	ADB90023	ADB90023 House com
12	8598	99.9	1663	7	ADD93518	ADD93518 Novel NOV
13	8598	99.9	1663	8	ADK12322	ADK12322 Human com
14	8598	99.9	1663	8	ADN04780	ADN04780 Antipsori
15	8597	99.9	1663	8	AAW40988	AAW40988 Human C3
16	8596	99.8	1663	2	AAW34621	AAW34621 Human C3
17	8596	99.8	1663	2	AAW34620	AAW34620 Human C3
18	8594	99.8	1663	2	AAW34029	AAW34029 Human mod
19	8593	99.8	1663	2	AAW34617	AAW34617 Human C3
20	8592	99.8	1667	2	AAW34631	AAW34631 Human C3
21	8588	99.8	1663	2	AAW34619	AAW34619 Human C3
22	8588	99.8	1663	2	AAW34628	AAW34628 Human C3
23	8586	99.7	1663	2	AAW34613	AAW34613 Human C3
24	8581	99.7	1663	2	AAW34030	AAW34030 Human mod
25	8581	99.7	1663	2	AAW34611	AAW34611 Human C3

26	8579	99.7	1663	2	AAW34614	Human C3
27	8578	99.6	1663	2	AAW34615	Human C3
28	8571	99.6	1663	2	AAW34612	Human C3
29	8567	99.5	1667	2	AAW34626	Human C3
30	8566	99.5	1663	2	AAW34630	Human C3
31	8554	99.4	1657	2	AAW34629	Human C3
32	8548	99.3	1663	2	AAW40990	Human C3
33	8545	99.3	1661	2	AAW34625	Human C3
34	8452	98.2	1635	2	AAW34624	Human C3
35	8193	95.2	1592	2	AAW34623	Human C3
36	7834.5	91.0	1540	4	ABG25976	Novel hum
37	6825	79.3	1663	8	ADI82104	Complemen
38	6752	78.4	1663	7	ADB90024	Mouse com
39	6752	78.4	1663	8	ADK12323	Mouse com
40	6272	72.9	1255	6	ADR63374	Human Alz
41	4386.5	51.0	1651	2	AAW63222	Pre-pro-c
42	4224	49.1	1642	2	AAW23729	Cobra ven
43	4224	49.1	1642	5	AAE21961	Pre-pro c
44	4224	49.1	1642	5	AAW71433	Cobra pro
45	4224	49.1	1650	5	AAW71469	Cobra pro

## ALIGNMENTS

RESULT 1  
AAW34606  
ID AAW34606 standard; protein; 1663 AA.  
XX  
AC AAW34606;  
XX  
DT 09-APR-1998 (first entry)  
XX  
DE Wild type human C3 protein.  
XX  
KW Human; C3 protein; convertase; complement pathway protein; infection;  
KW down-regulation resistant C3 convertase; xenograft rejection; therapy;  
KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;  
KW complement-mediated response; MHC-mismatched lymphocyte.  
XX  
OS Homo sapiens.  
XX  
PN WO9732981-A1.  
XX  
PD 12-SEP-1997.  
XX  
PF 04-MAR-1997; 97WO-GB000603.  
XX  
PR 07-MAR-1996; 96GB-00004865.  
PR 07-JUN-1996; 96GB-00011896.  
PR 08-JUL-1996; 96GB-00014293.  
PR 19-NOV-1996; 96GB-00024028.  
XX  
PA (IMUT-) IMUTRAN LTD.  
XX  
PI Farries TC, Harrison RA;  
XX  
PT WPI; 1997-457534/42.  
XX  
PS N-PSDB; AAT93749.  
XX  
PT Modified complement pathway protein that forms C3 convertase resistant to  
down-regulation - used to exhaust the complement pathway by super-  
activation, especially for preventing graft rejection, etc.  
XX  
PS Disclosure; Fig 1; 123pb; English.  
XX  
CC This sequence represents the wild type human C3 protein. This protein can  
be mutated to produce a protein of the invention. The protein of the  
invention is a modified native complement pathway protein (A) that forms  
a down-regulation resistant C3 convertase. (A), their variants, fragments  
CC and conjugates are used to deplete levels of complement pathway proteins  
(by superactivation until one or more components are exhausted),  
CC specifically to prevent rejection of foreign material (particularly a

CC xenograft) but also to prevent complement-mediated diseases resulting  
 CC from (surgical) injury or antibody-antigen interaction in autoimmune  
 CC disease, also to localise and/or amplify endogenous complement protein  
 CC conversion and deposition at a specific site (e.g. a virus, infected cell  
 CC or tumour, to increase sensitivity to complement-mediated responses; a  
 CC particular application is eliminating any cancer cells left after  
 CC surgical removal of a tumour). Also contemplated is ex vivo treatment,  
 CC especially by passing blood through a matrix containing (A) (this may  
 CC remove additional anaphylactic peptides and other inflammatory mediators)  
 CC or killing of leukaemia cells or MHC-mismatched lymphocytes in extracted  
 CC bone marrow. Since (A) is not inhibited by factor I, it can bind  
 CC repeatedly to factor B (which is then inactivated), causing inactivation  
 CC of the alternative pathway by consumption of factor B  
 XX

SQ Sequence 1663 AA;

Query Match 100.0%; Score 8609; DB 2; Length 1663;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPTSGPSLLLLTHLPLALGSPMYSITINILRSEETMVLBAHDAQDVPVTVTH 60  
 DB 1 MGPTSGPSLLLLTHLPLALGSPMYSITINILRSEETMVLBAHDAQDVPVTVTH 60  
 QY 61 DFGKKLVLSSSEKTLVTPATNHGNTFTIPANREFKSEKGRNFVTVQATFGTVQVEKV 120  
 DB 61 DFGKKLVLSSSEKTLVTPATNHGNTFTIPANREFKSEKGRNFVTVQATFGTVQVEKV 120  
 QY 121 VLVSQSGYLFQTDKTIYTPGSTVLYRIFVTHHKLPLVGTVMVNIENPEGIPVKQDSL 180  
 DB 121 VLVSQSGYLFQTDKTIYTPGSTVLYRIFVTHHKLPLVGTVMVNIENPEGIPVKQDSL 180  
 QY 181 SSQNLQGLPLSWDIPELVNMQWKIRAYENSPOQVFSTEFVEKVEYVLPSEFVIEPTE 240  
 DB 181 SSQNLQGLPLSWDIPELVNMQWKIRAYENSPOQVFSTEFVEKVEYVLPSEFVIEPTE 240  
 QY 241 KFYIYNKGLVETITARELYKKVEGTAFVIFGIDQGEQRIISLPESLKRIPEDSGSEV 300  
 DB 241 KFYIYNKGLVETITARELYKKVEGTAFVIFGIDQGEQRIISLPESLKRIPEDSGSEV 300  
 QY 301 VLSRKVLDDGVQNPRAEDLVGSLVSATVILHSGSDMVQAESGIPVTSPIQIHFTKT 360  
 DB 301 VLSRKVLDDGVQNPRAEDLVGSLVSATVILHSGSDMVQAESGIPVTSPIQIHFTKT 360  
 QY 361 PKYFKGMPFDLMVFTNPDGSPAYRVPVAVQGEDTVQSLTQGGVAKLSINTHPSQKPL 420  
 DB 361 PKYFKGMPFDLMVFTNPDGSPAYRVPVAVQGEDTVQSLTQGGVAKLSINTHPSQKPL 420  
 QY 421 SITVTRTKQELSEAEQATMTQALPYSIVGNSNNYVHLVRLTELPGETLVNVLRLMD 480  
 DB 421 SITVTRTKQELSEAEQATMTQALPYSIVGNSNNYVHLVRLTELPGETLVNVLRLMD 480  
 QY 481 RAHEAKIRYTYLIMNKGRLKAGQVREPQGLVPLSTITTPISPRVAVYTLIGA 540  
 DB 481 RAHEAKIRYTYLIMNKGRLKAGQVREPQGLVPLSTITTPISPRVAVYTLIGA 540  
 QY 541 SGQREVAVDSVWVDVDCVSGSLVVKSGQSDRQVPVQCGQMTLKEGDHGAARVVLVAVDK 600  
 DB 541 SGQREVAVDSVWVDVDCVSGSLVVKSGQSDRQVPVQCGQMTLKEGDHGAARVVLVAVDK 600  
 QY 601 GVEVLNKNKLTQSKITWDVVEKADIGCTPGSKDAGVFSADAGLTFTSSSQQTAQRAEL 660  
 DB 601 GVEVLNKNKLTQSKITWDVVEKADIGCTPGSKDAGVFSADAGLTFTSSSQQTAQRAEL 660  
 QY 661 QCQPAARRRRSVQLTEKEMDKVGYPKELRKCCEGDGMRENPMRFSQQRTRFISLGEAC 720  
 DB 661 QCQPAARRRRSVQLTEKEMDKVGYPKELRKCCEGDGMRENPMRFSQQRTRFISLGEAC 720  
 QY 721 KYVFLDCCNYITELRQHARASHGLGARNLDEDEIIAENIVSRSEPEWSLWNVDELKE 780  
 DB 721 KYVFLDCCNYITELRQHARASHGLGARNLDEDEIIAENIVSRSEPEWSLWNVDELKE 780  
 QY 781 PPKNGISTKLMNIFLKDSITTWELAVNSMSDKKICVADPFEVTVMQDFIDLELPSVYV 840

RESULT 2  
AAW34607

ID AAW34607 standard; protein; 1663 AA.

XX

AC AAW34607;

XX

DT 09-APR-1998 (first entry)

XX

DE Human C3 protein mutant C3M-123.

XX

KW Human; C3 protein; convertase; complement pathway protein; infection;

KW down-regulation resistant C3 convertase; xenograft rejection; therapy;

KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;

DB 781 PPKNGISTKLMNIFLKDSITTWELAVNSMSDKKICVADPFEVTVMQDFIDLELPSVYV 840  
 QY 841 RNEQVEIRAVLYNRYRONQELKVRVELLHNPAPCSLATTKRHCQTITPPKSSLSVPVVI 900  
 DB 841 RNEQVEIRAVLYNRYRONQELKVRVELLHNPAPCSLATTKRHCQTITPPKSSLSVPVVI 900  
 QY 901 VPLKTGLQVEVEKAAVYHHFISDGRKSLKVYVEGIRMMKTVAVRTLPERLREGVQKE 960  
 DB 901 VPLKTGLQVEVEKAAVYHHFISDGRKSLKVYVEGIRMMKTVAVRTLPERLREGVQKE 960  
 QY 961 DIPPADLSDDQVPTSETRILLOGTPVAQMTDAVDARELKLIVTPSGCGEQNMIGTTP 1020  
 DB 961 DIPPADLSDDQVPTSETRILLOGTPVAQMTDAVDARELKLIVTPSGCGEQNMIGTTP 1020  
 QY 1021 TVIAVHYLDETEQWERFLEKROGAELEIKKGYTQQLAFRQPSAFAAFVKRAFTWTLTA 1080  
 DB 1021 TVIAVHYLDETEQWERFLEKROGAELEIKKGYTQQLAFRQPSAFAAFVKRAFTWTLTA 1080  
 QY 1081 YVVKVPSLAVNLTAIDSOVLCCGAVKWLILEKOKPGVFOEDAPVTHQEWIIGLRNNKED 1140  
 DB 1081 YVVKVPSLAVNLTAIDSOVLCCGAVKWLILEKOKPGVFOEDAPVTHQEWIIGLRNNKED 1140  
 QY 1141 MALTAFLVILSLQBAKDICEEQVNSLPGSITKAGDFLEANNMQLORSYTVAIAGYALAQM 1200  
 DB 1141 MALTAFLVILSLQBAKDICEEQVNSLPGSITKAGDFLEANNMQLORSYTVAIAGYALAQM 1200  
 QY 1201 RLKGPLANKELTTAKOKNWEDEPGKOLYNVEATSVALLALLQLKDFDVPVVRWLNQOR 1260  
 DB 1201 RLKGPLANKELTTAKOKNWEDEPGKOLYNVEATSVALLALLQLKDFDVPVVRWLNQOR 1260  
 QY 1261 YGCGYGSTQATFWFQALAQYOKDAPDQELNVLVSLQPSRSKIITHRIHWESASLLR 1320  
 DB 1261 YGCGYGSTQATFWFQALAQYOKDAPDQELNVLVSLQPSRSKIITHRIHWESASLLR 1320  
 QY 1321 SEETKENGEGTVTAEGKGGTSLVNTMTHAKAKDQLTCKNFKDLKVTIKPAPETEKRPDA 1380  
 DB 1321 SEETKENGEGTVTAEGKGGTSLVNTMTHAKAKDQLTCKNFKDLKVTIKPAPETEKRPDA 1380  
 QY 1381 KNTMILBICTRYRGDDQDATMSILDISMTGFPAPDITDLKQLANGVDRIYSKYELDKAFSD 1440  
 DB 1381 KNTMILBICTRYRGDDQDATMSILDISMTGFPAPDITDLKQLANGVDRIYSKYELDKAFSD 1440  
 QY 1441 RNTLIIYLDKVSISEDCLAFKHQVFNVELIQPGAVKYVAYNLEESCTRFYHPEKEDG 1500  
 DB 1441 RNTLIIYLDKVSISEDCLAFKHQVFNVELIQPGAVKYVAYNLEESCTRFYHPEKEDG 1500  
 QY 1501 KLKLCRDELRCABENCFTQKSDDKVTLEERLDKACBEPGVYVYKTRLVKQLSNDPDE 1560  
 DB 1501 KLKLCRDELRCABENCFTQKSDDKVTLEERLDKACBEPGVYVYKTRLVKQLSNDPDE 1560  
 QY 1561 YMAIEQTIKSGDEVOVGQORFTFIPKICREALKLEEKHYLMGLSSDFWGEKPNLSY 1620  
 DB 1561 YMAIEQTIKSGDEVOVGQORFTFIPKICREALKLEEKHYLMGLSSDFWGEKPNLSY 1620  
 QY 1621 IIGKDTVWEHPBEDRCQDEENKQCDLGAFTESMVVFGCPN 1663  
 DB 1621 IIGKDTVWEHPBEDRCQDEENKQCDLGAFTESMVVFGCPN 1663

complement-mediated response; MHC-mismatched lymphocyte; mutein.

KW	complement-mediated response; MHC-mismatched lymphocyte; mutein.	QY	241	KFYIYNKGLVITITARELYGKKEGTAFVIFGIDGGEORISLPSISKRIPIEDSGEV	300
XX					
OS	Homo sapiens.	DB	241	KFYIYNKGLVITITARELYGKKEGTAFVIFGIDGGEORISLPSISKRIPIEDSGEV	300
XX					
FH	Key	QY	301	VLSRKVLGQVQNPRAEDLVGKSLYVSATVILHSGSDMVQAEKSGIPVTSYQIHFTKT	360
FT	Misc-difference 1303	DB	301	VLSRKVLGQVQNPRAEDLVGKSLYVSATVILHSGSDMVQAEKSGIPVTSYQIHFTKT	360
XX	/note= "Arg to Gln mutation"				
PN	WO9732981-A1.	QY	361	PKYFKGMPFDLMVFTNPDGSPAYRVPVAVOGEDVQSLTQGDGVAKLSINTHPSOKPL	420
XX					
PD	12-SEP-1997.	DB	361	PKYFKGMPFDLMVFTNPDGSPAYRVPVAVOGEDVQSLTQGDGVAKLSINTHPSOKPL	420
XX					
PF	04-MAR-1997; 97WO-GB0006503.	QY	421	SITVRTKQELSEAEQATRTMQALPYSTVGNNSNNVHLVSLRTELRLPGETNNVFLRMD	480
XX					
PR	07-MAR-1996; 96GB-00004865.	DB	421	SITVRTKQELSEAEQATRTMQALPYSTVGNNSNNVHLVSLRTELRLPGETNNVFLRMD	480
PR	07-JUN-1996; 96GB-00011896.	QY	481	RAHEAKIRYVYVILMNKGRLLKAGQVREPQGLVVLPLSLITTFPIPSFRVAVYVTLGA	540
PR	08-JUL-1996; 96GB-00014293.	DB	481	RAHEAKIRYVYVILMNKGRLLKAGQVREPQGLVVLPLSLITTFPIPSFRVAVYVTLGA	540
PR	19-NOV-1996; 96GB-00024028.	QY	541	SGOREVVADSVVVDKSDCVGSLVVKSGQSEDRQVPQGMVTLKIEGDHGRVVLVAVDK	600
XX	(IMUT-) IMUTRAN LTD.	DB	541	SGOREVVADSVVVDKSDCVGSLVVKSGQSEDRQVPQGMVTLKIEGDHGRVVLVAVDK	600
XX					
PI	Farries TC, Harrison RA;	QY	601	GVFLNKNKLTQSKINDVVEKADIGCTPGSKDYAGVFSADGLTFTSSGQCTAQAEL	660
XX					
DR	WPI; 1997-457534/42.	DB	601	GVFLNKNKLTQSKINDVVEKADIGCTPGSKDYAGVFSADGLTFTSSGQCTAQAEL	660
XX					
PT	Modified complement pathway protein that forms C3 convertase resistant to	QY	661	QCPQPAARRRRSVOLTEKMDKVKYKPELRLKCCSDGMRNPMRPFSCORRTFISLGEAC	720
PT	down-regulation - used to exhaust the complement pathway by super-	DB	661	QCPQPAARRRRSVOLTEKMDKVKYKPELRLKCCSDGMRNPMRPFSCORRTFISLGEAC	720
PT	activation, especially for preventing graft rejection, etc.	QY	721	KVFLDCCNVITELRQAHARASHGLARSLNLDIEDIIAENIVSRSEFPESLWNVDELKE	780
XX					
PS	Example 4; Page; 123pp; English.	DB	721	KVFLDCCNVITELRQAHARASHGLARSLNLDIEDIIAENIVSRSEFPESLWNVDELKE	780
XX					
CC	This sequence represents a mutated human C3 protein of the invention (see	QY	781	PPKNGISTKMLNIFLKDSTITWEILAVSMDSKKGICVADPPFEVTVMDQFFIDLRLPYGVV	840
CC	AAW34606 for wild type protein). This protein is a protein of the	DB	781	PPKNGISTKMLNIFLKDSTITWEILAVSMDSKKGICVADPPFEVTVMDQFFIDLRLPYGVV	840
CC	invention, and is a modified native complement pathway protein (A) that	QY	841	RNEQVEIRAVLYNRQONQELKVRVLLHNPAFCSLATTKRRHQQTITIPPKSSLSVPVVI	900
CC	forms a down-regulation resistant C3 convertase. (A), their variants,	DB	841	RNEQVEIRAVLYNRQONQELKVRVLLHNPAFCSLATTKRRHQQTITIPPKSSLSVPVVI	900
CC	fragments and conjugates are used to deplete levels of complement pathway	QY	901	VPLKGLQEVKAAVYHHFISDGVKSLKVVPEGRMKNVAVRTLDPERLGRGVQKE	960
CC	proteins (by superactivation until one or more components are exhausted),	DB	901	VPLKGLQEVKAAVYHHFISDGVKSLKVVPEGRMKNVAVRTLDPERLGRGVQKE	960
CC	specifically to prevent rejection of foreign material (particularly a	QY	961	DIPPADLSQVDPDTESETRILLQGTVAQMTEDAVDAERLKHILVTPSGCGEONMIGMTP	1020
CC	xenograft) but also to prevent complement-mediated diseases resulting	DB	961	DIPPADLSQVDPDTESETRILLQGTVAQMTEDAVDAERLKHILVTPSGCGEONMIGMTP	1020
CC	from (surgical) injury or antibody-antigen interaction in autoimmune	QY	1021	TVIAVHYLDETEQWEKFGLEKQGALELIIKGYTQQLAFRQPSFAFAFVKRAPSTWLTATA	1080
CC	disease, also to localise and/or amplify endogenous complement protein	DB	1021	TVIAVHYLDETEQWEKFGLEKQGALELIIKGYTQQLAFRQPSFAFAFVKRAPSTWLTATA	1080
CC	conversion and deposition at a specific site (e.g. a virus, infected cell	QY	1081	VYVKVPSLANVLIADISQVLCGAVKWLILEKQKPGVFCEDAPVHQEMIGLRRNNEKD	1140
CC	or tumour, to increase sensitivity to complement-mediated responses; a	DB	1081	VYVKVPSLANVLIADISQVLCGAVKWLILEKQKPGVFCEDAPVHQEMIGLRRNNEKD	1140
CC	particular application is eliminating any cancer cells left after	QY	1141	MALTAFVLISLQAKDICEQVNSLPGSITKAGDFLEANYMNLQRSYTVIAGYALAQMG	1200
CC	surgical removal of a tumour). Also contemplated is ex vivo treatment,	DB	1141	MALTAFVLISLQAKDICEQVNSLPGSITKAGDFLEANYMNLQRSYTVIAGYALAQMG	1200
CC	especially by passing blood through a matrix containing (A) (this may	QY	1201	RLKGPLLNKFLTTAKDKRNEWDPGKQLYVNEATSVALLALLQKDFDVPVFWVWLNQOR	1260
CC	remove additional anaphylactic peptides and other inflammatory mediators)	DB	1201	RLKGPLLNKFLTTAKDKRNEWDPGKQLYVNEATSVALLALLQKDFDVPVFWVWLNQOR	1260
CC	or killing of leukaemia cells or MHC-mismatched lymphocytes in extracted	QY	1261	YGGGYSQVSTQATFVNFQALAQYOKDAPDHOELNLDVSLQLPSRSKITHRIHWESASLLR	1320
CC	bone marrow. Since (A) is not inhibited by factor I, it can bind	DB	1261	YGGGYSQVSTQATFVNFQALAQYOKDAPDHOELNLDVSLQLPSRSKITHRIHWESASLLR	1320
CC	repeatedly to factor B (which is then inactivated), causing inactivation				
CC	of the alternative pathway by consumption of factor B				
XX					
SQ	Sequence 1663 AA;				

Query Match 100.0%; Score 8605; DB 2; Length 1663;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1662; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGPTSGPSLLLLTHLPLALGSPMSIITPNILRLSEBETVYLEAHDAGQDVPVTVVH	60
DB	1	MGPTSGPSLLLLTHLPLALGSPMSIITPNILRLSEBETVYLEAHDAGQDVPVTVVH	60
QY	61	DFPGKLVLSSEKTVLTPATNHGNTFTTIPANREFSKGRNKFVTVQATFGTVQVKV	120
DB	61	DFPGKLVLSSEKTVLTPATNHGNTFTTIPANREFSKGRNKFVTVQATFGTVQVKV	120
QY	121	VLSLQSGYLFIQDKTIYTPGSTVLYRIFTVNHKLLPVGRVWNIENPEGIPVKQDSL	180
DB	121	VLSLQSGYLFIQDKTIYTPGSTVLYRIFTVNHKLLPVGRVWNIENPEGIPVKQDSL	180
QY	181	SSQNLQGLVPLSWDIPELVNMGQMKRAYYENSQQVSTEFVEKBYVLPSEVIVEPTE	240
DB	181	SSQNLQGLVPLSWDIPELVNMGQMKRAYYENSQQVSTEFVEKBYVLPSEVIVEPTE	240

QY 1321 SEETKENEGFTVTAEGKGOTLSVVTMHHAKADOLTCNKFDLKVTKIPAPETEKRPQDA 1380  
 DB 1321 SEETKENEGFTVTAEGKGOTLSVVTMHHAKADOLTCNKFDLKVTKIPAPETEKRPQDA 1380  
 QY 1381 KNTMILEICTRVGGDDATMSILDISMGTGFPAPDITDLDKOLANGVDYVYISKYELDKAFSD 1440  
 DB 1381 KNTMILEICTRVGGDDATMSILDISMGTGFPAPDITDLDKOLANGVDYVYISKYELDKAFSD 1440  
 QY 1441 RNTLIITVLDKSHSEDDCLAFKQHVFNVELIQPGAVKYAYXNLEESCTRFYHPEKEDG 1500  
 DB 1441 RNTLIITVLDKSHSEDDCLAFKQHVFNVELIQPGAVKYAYXNLEESCTRFYHPEKEDG 1500  
 QY 1501 KLNKLRDELRCRAENCFIQSDDKVTLBEERLDKACFPGVDYVYKPLVKVQLSNDFDE 1560  
 DB 1501 KLNKLRDELRCRAENCFIQSDDKVTLBEERLDKACFPGVDYVYKPLVKVQLSNDFDE 1560  
 QY 1561 YIMAIETIKSGSDEYVQVQQRFTFSPKICREALKLEBKHYLMWGLSSDFWGEKPNLSY 1620  
 DB 1561 YIMAIETIKSGSDEYVQVQQRFTFSPKICREALKLEBKHYLMWGLSSDFWGEKPNLSY 1620  
 QY 1621 IIGKDTWVHWPEDSCDEENKQKQODLGAFTESMWVFGCPN 1663  
 DB 1621 IIGKDTWVHWPEDSCDEENKQKQODLGAFTESMWVFGCPN 1663

## RESULT 3

AAW34610  
 ID AAW34610 standard; protein; 1663 AA.

AC AAW34610;

DT 09-APR-1998 (first entry)

DE Human C3 protein mutant CV-2.

XX Human: C3 protein; convertase; complement pathway protein; infection;  
 KW down-regulation resistant C3 convertase; xenograft rejection; therapy;  
 KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;  
 KW complement-mediated response; MHC-mismatched lymphocyte; muten.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 776 /note= "E776K mutation"

XX

XX W09732981-A1.

XX 12-SEP-1997.

XX

XX 04-MAR-1997; 97MO-GB000603.

XX

XX 07-MAR-1996; 96GB-00004865.

XX 07-JUN-1996; 96GB-00011896.

XX 08-JUL-1996; 96GB-00014293.

XX 19-NOV-1996; 96GB-00024028.

XX

XX (IMUT-) IMUTRAN LTD.

XX

XX Farries TC, Harrison RA;

XX WPI; 1997-457534/42.

XX

XX Modified complement pathway protein that forms C3 convertase resistant to

XX down-regulation - used to exhaust the complement pathway by super-

XX activation, especially for preventing graft rejection, etc.

CC fragments and conjugates are used to deplete levels of complement pathway  
 CC proteins (by superactivation until one or more components are exhausted),  
 CC specifically to prevent rejection of foreign material (particularly a  
 CC xenograft) but also to prevent complement-mediated diseases resulting  
 CC from (surgical) injury or antibody-antigen interaction in autoimmune  
 CC disease, also to localise and/or amplify endogenous complement protein  
 CC conversion and deposition at a specific site (e.g. a virus, infected cell  
 CC or tumour, to increase sensitivity to complement-mediated responses; a  
 CC particular application is eliminating any cancer cells left after  
 CC surgical removal of a tumour). Also contemplated is ex vivo treatment,  
 CC especially by passing blood through a matrix containing (A) (this may  
 CC remove additional anaphylactic peptides and other inflammatory mediators)  
 CC or killing of leukaemia cells or MHC-mismatched lymphocytes in extracted  
 CC bone marrow. Since (A) is not inhibited by factor 1, it can bind  
 CC repeatedly to factor B (which is then inactivated), causing inactivation  
 CC of the alternative pathway by consumption of factor B  
 XX  
 SQ Sequence 1663 AA;

Query Match 100.0%; Score 8605; DB 2; Length 1663;

Best Local Similarity 99.9%; Pred No 0;

Matches 1662; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPTSGPSLLLLTHLPALGSPMYSIITPNILRLESBETMWLEAHDAGQDVPVTVVH 60

DB 1 MGPTSGPSLLLLTHLPALGSPMYSIITPNILRLESBETMWLEAHDAGQDVPVTVVH 60

QY 61 DPGKKLVLSSEKTVLTPATNMGNTVFTIPANREFKSEKGRNKFVTVQATFGTOVEKV 120

DB 61 DPGKKLVLSSEKTVLTPATNMGNTVFTIPANREFKSEKGRNKFVTVQATFGTOVEKV 120

QY 121 VLVSQSGYLFQTDKTYITFGSTVLYRIPTVNHKLLPVGRITVMNENPEGIPVKQDSL 180

DB 121 VLVSQSGYLFQTDKTYITFGSTVLYRIPTVNHKLLPVGRITVMNENPEGIPVKQDSL 180

QY 181 SSKQNLGVLPLSWDIPBLVNMGWKIRAYVENSPPQVFSTEPKEVYLPSEVIVEPTE 240

DB 181 SSKQNLGVLPLSWDIPBLVNMGWKIRAYVENSPPQVFSTEPKEVYLPSEVIVEPTE 240

QY 241 KFYIYNEKGLVITITARFLYKKGVECTAPVIFGIGQSGQRISLPESLKRIPIEGSGEV 300

DB 241 KFYIYNEKGLVITITARFLYKKGVECTAPVIFGIGQSGQRISLPESLKRIPIEGSGEV 300

QY 301 VLSRKVLDDGVQNPRAEDLVGKSLYVSATVILHSGSDMWQAERSGPIVTSFYQHFTKT 360

DB 301 VLSRKVLDDGVQNPRAEDLVGKSLYVSATVILHSGSDMWQAERSGPIVTSFYQHFTKT 360

QY 361 PRYFKPGMPFDLMVFTVNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHPSQKPL 420

DB 361 PRYFKPGMPFDLMVFTVNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHPSQKPL 420

QY 421 SITVTRKKOELSEAEQATRTMQLPYSTVGNNSNNYLHLSVLRTLPGETLNVNFIIRMD 480

DB 421 SITVTRKKOELSEAEQATRTMQLPYSTVGNNSNNYLHLSVLRTLPGETLNVNFIIRMD 480

QY 481 RAHEAKIRYTYLIMNKGRLIKAGQVRBPQDLVPLSIITTDFTPSRLVAYTYLIGA 540

DB 481 RAHEAKIRYTYLIMNKGRLIKAGQVRBPQDLVPLSIITTDFTPSRLVAYTYLIGA 540

QY 541 SGOREVADSVWVDVKDSCVGSVVKSGQSEDRQVPFGQMTLKIEGDHGARVAVDXK 600

DB 541 SGOREVADSVWVDVKDSCVGSVVKSGQSEDRQVPFGQMTLKIEGDHGARVAVDXK 600

QY 601 GVFLVKNKVKLTQSKINDVVEKADIGCTPGSGDKYAGVFSAGLTITSSSGQQTARABL 660

DB 601 GVFLVKNKVKLTQSKINDVVEKADIGCTPGSGDKYAGVFSAGLTITSSSGQQTARABL 660

QY 661 QCPOPAARRRRSVQITKRMVKVGPKELRKCCEDGMRENPMRPFSCQRTTFISLGEAC 720

DB 661 QCPOPAARRRRSVQITKRMVKVGPKELRKCCEDGMRENPMRPFSCQRTTFISLGEAC 720

QY 721 KKVFLDCCNYITELAROHARASHLGLARSLNDEDTIAENIVSRSEFPSSMWNVEDLKE 780

DB 721 KKVFLDCCNYITELAROHARASHLGLARSLNDEDTIAENIVSRSEFPSSMWNVEDLKE 780

Db 721 KKVFLDCCNYITELRQHRARASHGLARSNLDIEDIIAEENIVSRSEFPBESLWLVNVDLKE 780  
 Qy 781 PPKNGISLKNWIFLKDSITTWELAVMSDKKIGICVADPEVTVQDFFIDRLRLPSYV 840  
 Db 781 PPKNGISLKNWIFLKDSITTWELAVMSDKKIGICVADPEVTVQDFFIDRLRLPSYV 840  
 Qy 841 RNEQVEIRAVLYNRQNELKVRVELLHNFAPCSLATTKRHHQQTITIPKSSLSVPYVI 900  
 Db 841 RNEQVEIRAVLYNRQNELKVRVELLHNFAPCSLATTKRHHQQTITIPKSSLSVPYVI 900  
 Qy 901 VPLKTGLQEVSKAAVYHFFISDGVKSLKVPBGIRMNKTAVRTLDPERLREGVQKE 960  
 Db 901 VPLKTGLQEVSKAAVYHFFISDGVKSLKVPBGIRMNKTAVRTLDPERLREGVQKE 960  
 Qy 961 DIPPADLSQVDPDTESETRILLQGTTPVAQMTEDAVDAERLKLHIVTPSGCGEQNMIGMTP 1020  
 Db 961 DIPPADLSQVDPDTESETRILLQGTTPVAQMTEDAVDAERLKLHIVTPSGCGEQNMIGMTP 1020  
 Qy 1021 TVIAVHYLDETEQWEKLEKROGALBELIKGYTOOLAFPOSSAFAAFVKRAPSFWLTA 1080  
 Db 1021 TVIAVHYLDETEQWEKLEKROGALBELIKGYTOOLAFPOSSAFAAFVKRAPSFWLTA 1080  
 Qy 1081 YVVKVFSIAVNLIAIDSOVLGCAVKNLILEKQKPDGVFORDAFVHQQEMIGLNNNEKD 1140  
 Db 1081 YVVKVFSIAVNLIAIDSOVLGCAVKNLILEKQKPDGVFORDAFVHQQEMIGLNNNEKD 1140  
 Qy 1141 MALTAFLVLSIQEAKDICEQVNSLPGSITKAGDFLEANTYMLQRTVTAIAGYALAOAG 1200  
 Db 1141 MALTAFLVLSIQEAKDICEQVNSLPGSITKAGDFLEANTYMLQRTVTAIAGYALAOAG 1200  
 Qy 1201 RLKGPLNKFITTTAKDNKRWEDPGQLYNVEATSYALLALLQKDFDPPVVRWLNQOR 1260  
 Db 1201 RLKGPLNKFITTTAKDNKRWEDPGQLYNVEATSYALLALLQKDFDPPVVRWLNQOR 1260  
 Qy 1261 YVGGYGSTQATEVVFQALQYQKADPDHOLNLDVSLQPSSEKITHRIWESASLAR 1320  
 Db 1261 YVGGYGSTQATEVVFQALQYQKADPDHOLNLDVSLQPSSEKITHRIWESASLAR 1320  
 Qy 1321 SEETKENEFTVTAEGKGGQGLSVVWYHAKAKDQLTCKNFKDLKVTIKPAPETERPQDA 1380  
 Db 1321 SEETKENEFTVTAEGKGGQGLSVVWYHAKAKDQLTCKNFKDLKVTIKPAPETERPQDA 1380  
 Qy 1381 KNTWILEICTRYRGDQATWSILDISMTGFAPDTPDLDKQANGVDVRYISKYELDKAFSD 1440  
 Db 1381 KNTWILEICTRYRGDQATWSILDISMTGFAPDTPDLDKQANGVDVRYISKYELDKAFSD 1440  
 Qy 1441 RNTLIILYDKVSHSEDDCLAPKHQYFNVELIQGAVKVVAYVYNLEESCTRFYHPEKEDG 1500  
 Db 1441 RNTLIILYDKVSHSEDDCLAPKHQYFNVELIQGAVKVVAYVYNLEESCTRFYHPEKEDG 1500  
 Qy 1501 KLNKL-CRDEL-CRAEENCFIQSDDKVTLERLDKACEPGVDVYVYKTRLVKQLSNDPDE 1560  
 Db 1501 KLNKL-CRDEL-CRAEENCFIQSDDKVTLERLDKACEPGVDVYVYKTRLVKQLSNDPDE 1560  
 Qy 1561 YMAIEQTIKSGSDEVQVQOQRTFISPKCREALKLEKHYLMWGLSDFWGEKPNLSY 1620  
 Db 1561 YMAIEQTIKSGSDEVQVQOQRTFISPKCREALKLEKHYLMWGLSDFWGEKPNLSY 1620  
 Qy 1621 IIGKDTWVHFWPEDECODENQOCODLGAFTESVVFQCPN 1663  
 Db 1621 IIGKDTWVHFWPEDECODENQOCODLGAFTESVVFQCPN 1663  
 RESULT 4  
 ID AAW34609  
 AC AAW34609 standard; protein; 1663 AA.  
 XX AAW34609;  
 DT 09-APR-1998 (first entry)  
 XX Human C3 protein mutant C3M-51.  
 XX

KW Human: C3 protein; convertase; complement pathway protein; infection;  
 KW down-regulation resistant C3 convertase; xenograft rejection; therapy;  
 KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;  
 KW complement-mediated response; MHC-mismatched lymphocyte; mutein.  
 OS Homo sapiens.  
 XX  
 Key Location/Qualifiers  
 Misc-difference 1320  
 /note= "Arg to Gln mutation"  
 WT W09732981-A1.  
 PD 12-SEP-1997.  
 PF 04-MAR-1997; 97MO-GB000603.  
 PR 07-MAR-1996; 96GB-00004865.  
 PR 07-JUN-1996; 96GB-00011896.  
 PR 08-JUL-1996; 96GB-00014293.  
 PR 19-NOV-1996; 96GB-00024028.  
 XX (IMUT-) IMUTRAN LTD.  
 PA Farries TC, Harrison RA;  
 PI WPI; 1997-457534/42.  
 XX  
 DR Modified complement pathway protein that forms C3 convertase resistant to  
 XX down-regulation - used to exhaust the complement pathway by super-  
 PT activation, especially for preventing graft rejection, etc.  
 PT  
 PS Claim 7; Page; 123pp; English.  
 XX

CC This sequence represents a mutated human C3 protein of the invention (see  
 CC AAW34609 for wild type protein). This protein is a protein of the  
 CC invention, and is a modified native complement pathway protein (A) that  
 CC forms a down-regulation resistant C3 convertase. (A), their variants,  
 CC fragments and conjugates are used to deplete levels of complement pathway  
 CC proteins (by superactivation until one or more components are exhausted),  
 CC specifically to prevent rejection of foreign material (particularly a  
 CC xenograft) but also to prevent complement-mediated diseases resulting  
 CC from (surgical) injury or antibody-antigen interaction in autoimmune  
 CC disease, also to localise and/or amplify endogenous complement protein  
 CC conversion and deposition at a specific site (e.g. a virus, infected cell  
 CC or tumour, to increase sensitivity to complement-mediated responses; a  
 CC particular application is eliminating any cancer cells left after  
 CC surgical removal of a tumour). Also contemplated is ex vivo treatment,  
 CC especially by passing blood through a matrix containing (A) (this may  
 CC remove additional anaphylactic peptides and other inflammatory mediators)  
 CC or killing of leukaemia cells or MHC-mismatched lymphocytes in extracted  
 CC bone marrow. Since (A) is not inhibited by factor I, it can bind  
 CC repeatedly to factor B (which is then inactivated), causing inactivation  
 CC of the alternative pathway by consumption of factor B  
 XX

SQ Sequence 1663 AA;  
 Query Match 100.0%; Score 8605; DB 2; Length 1663;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1662; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MGPTSGPSLLLLTHLPALGSPMYSIITPNILRESEETMVLSEAHDAQSDVPVTVVH 60  
 Db 1 MGPTSGPSLLLLTHLPALGSPMYSIITPNILRESEETMVLSEAHDAQSDVPVTVVH 60  
 Qy 61 DFFGKLVLSSEKTVLTPATNHMGVNTFTIPANREFKSEKGNKFTVQATFGTVVEKV 120  
 Db 61 DFFGKLVLSSEKTVLTPATNHMGVNTFTIPANREFKSEKGNKFTVQATFGTVVEKV 120  
 Qy 121 VLVSQSGVLFQTDKTIYTPGSTVLYRIFTVNHKLLPVGRVWVNIENPEGIPVKQDSL 180  
 Db 121 VLVSQSGVLFQTDKTIYTPGSTVLYRIFTVNHKLLPVGRVWVNIENPEGIPVKQDSL 180

QY 181 SSQNLGVLPLSDIPELVNMGWKIRAYVENSPOQVFSTPEVKEVYLPSFEVIVPE 240  
 Db 181 SSQNLGVLPLSDIPELVNMGWKIRAYVENSPOQVFSTPEVKEVYLPSFEVIVPE 240  
 QY 241 KFYIYNEKLEVTITARFLYKKVGTARFIQDGEORISLPESLKRPIEDGSGEV 300  
 Db 241 KFYIYNEKLEVTITARFLYKKVGTARFIQDGEORISLPESLKRPIEDGSGEV 300  
 QY 301 VLSRKVLIDGQNPRAEDLVKSLYSATVILHSGSDMWQAERSGIPVTSFYQHFTKT 360  
 Db 301 VLSRKVLIDGQNPRAEDLVKSLYSATVILHSGSDMWQAERSGIPVTSFYQHFTKT 360  
 QY 361 KYKFGMPDLVFNTPDGSFAYRVPVAVQGEDTVQSLTQDGDVAKLSINTHPSQKPL 420  
 Db 361 KYKFGMPDLVFNTPDGSFAYRVPVAVQGEDTVQSLTQDGDVAKLSINTHPSQKPL 420  
 QY 421 SITVTRTKQELSEAEQATRTMQALPYSTVGNSSNNYLHLSVLRTLPGETLNVPFLRMD 480  
 Db 421 SITVTRTKQELSEAEQATRTMQALPYSTVGNSSNNYLHLSVLRTLPGETLNVPFLRMD 480  
 QY 481 RAHAKIRYTYILNMKGRLLKAGORVRSPGDVLPVLSITTDITPSFRLVAYYTLGA 540  
 Db 481 RAHAKIRYTYILNMKGRLLKAGORVRSPGDVLPVLSITTDITPSFRLVAYYTLGA 540  
 QY 541 SGOREVWADSVWVDVKDSCVGLSVKSGQSEDRQPVPGQOMTLKIEGSHGARVVLVADK 600  
 Db 541 SGOREVWADSVWVDVKDSCVGLSVKSGQSEDRQPVPGQOMTLKIEGSHGARVVLVADK 600  
 QY 601 GFVLNKNKLTKQSKINDVVEKADIGCTPGSGKDYAGVPSDAGLTFTSSSGOQRAEL 660  
 Db 601 GFVLNKNKLTKQSKINDVVEKADIGCTPGSGKDYAGVPSDAGLTFTSSSGOQRAEL 660  
 QY 661 QCPQPAARRRSVOLTEKRMKVKYKPKSLRKCCEDEGMENPMRPSQORRTFISLGEAC 720  
 Db 661 QCPQPAARRRSVOLTEKRMKVKYKPKSLRKCCEDEGMENPMRPSQORRTFISLGEAC 720  
 QY 721 KKFVLDCCNITELRRORHARASHGLARSLNDEDIIAENIVSRSEFPESLWNVEDLKE 780  
 Db 721 KKFVLDCCNITELRRORHARASHGLARSLNDEDIIAENIVSRSEFPESLWNVEDLKE 780  
 QY 781 PKKNGISTKLMNIFLKDSTITWILAVMSDKKICVADPPVTVNQOFFIDILPLPSVY 840  
 Db 781 PKKNGISTKLMNIFLKDSTITWILAVMSDKKICVADPPVTVNQOFFIDILPLPSVY 840  
 QY 841 RNEQVEIRAVLYNRQNELKVRVELLHNPFAFCSLATTKRHQQTITIPPKSSLSVPYVI 900  
 Db 841 RNEQVEIRAVLYNRQNELKVRVELLHNPFAFCSLATTKRHQQTITIPPKSSLSVPYVI 900  
 QY 901 VPLKTGLQEVKAAVYHHFISDGVKRSIKVYPEGIRMNKTAVATLDPERLREGVQKE 960  
 Db 901 VPLKTGLQEVKAAVYHHFISDGVKRSIKVYPEGIRMNKTAVATLDPERLREGVQKE 960  
 QY 961 DIPPADLSDOVPDTESETRILLQGTPEVAQMTEDAVDAERLKHILVTPSGCGEQNMIGWTP 1020  
 Db 961 DIPPADLSDOVPDTESETRILLQGTPEVAQMTEDAVDAERLKHILVTPSGCGEQNMIGWTP 1020  
 QY 1021 TVIAVHVLDETEQWEKFGLEKROGALELTKGYTQCLAPROPSSAFAPVKGASTWLTA 1080  
 Db 1021 TVIAVHVLDETEQWEKFGLEKROGALELTKGYTQCLAPROPSSAFAPVKGASTWLTA 1080  
 QY 1081 YVVKVFLAVNLIAIDSVQLCGAVKWLILLEKQKPGVQEDAPVIRQEMIGGLRNNNEKD 1140  
 Db 1081 YVVKVFLAVNLIAIDSVQLCGAVKWLILLEKQKPGVQEDAPVIRQEMIGGLRNNNEKD 1140  
 QY 1141 MALTAFLVLSIQAKDICEBOVNSLPGSITKAGDFLEANYNMLQBSYTVATAGYALQMG 1200  
 Db 1141 MALTAFLVLSIQAKDICEBOVNSLPGSITKAGDFLEANYNMLQBSYTVATAGYALQMG 1200  
 QY 1201 RLKGFPLNKFLTTAKOKRWEDSGKOLYNVEATSVALLALQLKDFDFVPPVFWLNEQR 1260  
 Db 1201 RLKGFPLNKFLTTAKOKRWEDSGKOLYNVEATSVALLALQLKDFDFVPPVFWLNEQR 1260  
 QY 1261 YVGGYGSTQATPMVFWQALAQYKQDAPDHQELNLDVSLQLPSSSKIITHRWESASLLR 1320

Db 1261 YVGGYGSTQATPMVFWQALAQYKQDAPDHQELNLDVSLQLPSSSKIITHRWESASLLR 1320  
 QY 1321 SEETKENGFTVTAEGKGQGTLSVVTMYHAKAQDQLTCKNFKDLKYTIKAPETEKRPQDA 1380  
 Db 1321 SEETKENGFTVTAEGKGQGTLSVVTMYHAKAQDQLTCKNFKDLKYTIKAPETEKRPQDA 1380  
 QY 1381 KVTMLLEICTRYRGDQDATMSILDISMVTFAPDPTDQLQLANGVDVRYISKYELDKAFSD 1440  
 Db 1381 KVTMLLEICTRYRGDQDATMSILDISMVTFAPDPTDQLQLANGVDVRYISKYELDKAFSD 1440  
 QY 1441 RNTLIIVLDKYSHSDDECLAFKHQVFNVELIQPGAVKYVAYYNLEESCTRYHBEKSDG 1500  
 Db 1441 RNTLIIVLDKYSHSDDECLAFKHQVFNVELIQPGAVKYVAYYNLEESCTRYHBEKSDG 1500  
 QY 1501 KLNLCDRELCDRCABENCFIQKSDDKVTLLEERLDKACFEGVDYVYKTRLVKQVLSNDFDE 1560  
 Db 1501 KLNLCDRELCDRCABENCFIQKSDDKVTLLEERLDKACFEGVDYVYKTRLVKQVLSNDFDE 1560  
 QY 1561 YMALEQTIKSGSDEVQVGOORTFISPIKCREALKLEEKHYLMGLSSDFWGEKPNLSY 1620  
 Db 1561 YMALEQTIKSGSDEVQVGOORTFISPIKCREALKLEEKHYLMGLSSDFWGEKPNLSY 1620  
 QY 1621 IIGKDTWVEHWPEDBECQDBENQKQCDLGAFTESWVFGCPN 1663  
 Db 1621 IIGKDTWVEHWPEDBECQDBENQKQCDLGAFTESWVFGCPN 1663  
 RESULT 5  
 AAW34616  
 ID AAW34616 standard; protein; 1663 AA.  
 XX AAW34616;  
 AC AC  
 DT 09-APR-1998 (first entry)  
 XX Human C3 protein mutant CV-5.  
 DE Human; C3 protein; convertase; complement pathway protein; infection;  
 KW down-regulation resistant C3 convertase; xenograft rejection; therapy;  
 KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;  
 XX complement-mediated response; MHC-mismatched lymphocyte; mutein.  
 OS Homo sapiens.  
 Key Location/Qualifiers  
 EH Misc-difference 1134  
 FT /note= "R1134Q mutation"  
 XX WO9732981-A1.  
 PN 12-SEP-1997.  
 PD 04-MAR-1997; 97WO-GB000603.  
 PF 07-MAR-1996; 96GB-00004865.  
 PR 07-JUN-1996; 96GB-00011896.  
 PR 08-JUL-1996; 96GB-00014293.  
 PR 19-NOV-1996; 96GB-00024028.  
 XX (IMUT-) IMUTRAN LTD.  
 PA Farries TC, Harrison RA;  
 XX WPI; 1997-457534/42.  
 XX Modified complement pathway protein that forms C3 convertase resistant to  
 PT down-regulation - used to exhaust the complement pathway by super-  
 PS activation, especially for preventing graft rejection, etc.  
 XX Example 14; Page; 123pp; English.  
 XX This sequence represents a mutated human C3 protein of the invention (see



AAW34606 for wild type protein). This protein is a protein of the invention, and is a modified native complement pathway protein (A) that forms a down-regulation resistant C3 convertase. (A), their variants, fragments and conjugates are used to deplete levels of complement pathway proteins (by superactivation until one or more components are exhausted), specifically to prevent rejection of foreign material (particularly a xenograft) but also to prevent complement-mediated diseases resulting from (surgical) injury or antibody-antigen interaction in autoimmune disease, also to localise and/or amplify endogenous complement protein conversion and deposition at a specific site (e.g. a virus, infected cell or tumour, to increase sensitivity to complement-mediated responses; a particular application is eliminating any cancer cells left after surgical removal of a tumour). Also contemplated is ex vivo treatment, especially by passing blood through a matrix containing (A) (this may remove additional anaphylactic peptides and other inflammatory mediators) or killing of leukaemia cells or MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not inhibited by factor I, it can bind repeatedly to factor B (which is then inactivated), causing inactivation of the alternative pathway by consumption of factor B.

Sequence 1663 AA;

Query Match 100.0%; Score 8605; DB 2; Length 1663;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1662; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPTSGPSLLILLTHLPLALGSPMYSIIIPNLRLESEBTMVLHAHQDQGVPTVTVH 60  
DB 1 MGPTSGPSLLILLTHLPLALGSPMYSIIIPNLRLESEBTMVLHAHQDQGVPTVTVH 60  
QY 61 DFPQKVLVSSEKTVLTPATNMGNVFTTIPANREFKSEKGRNKFVTVQATFGTQVVEKV 120  
DB 61 DFPQKVLVSSEKTVLTPATNMGNVFTTIPANREFKSEKGRNKFVTVQATFGTQVVEKV 120  
QY 121 VLVSLSQGYLFIOFTDKTYIPGSTVLYRIETVNHKLLPVGRVTVMNIENPEGIPVKQDSL 180  
DB 121 VLVSLSQGYLFIOFTDKTYIPGSTVLYRIETVNHKLLPVGRVTVMNIENPEGIPVKQDSL 180  
QY 181 SSQQLGVLPDLSWDIPELVNMQWKIRAYYENSPPQVFSTEFVEKYEVLPSFEVIVEPTE 240  
DB 181 SSQQLGVLPDLSWDIPELVNMQWKIRAYYENSPPQVFSTEFVEKYEVLPSFEVIVEPTE 240  
QY 241 KFYIYNEKGLGVITITARELYGKVEGTAFVIGIODEGQRIISLPSLXRIEIDGSGEV 300  
DB 241 KFYIYNEKGLGVITITARELYGKVEGTAFVIGIODEGQRIISLPSLXRIEIDGSGEV 300  
QY 301 VLSRKVLDDGQNPRAEDLVGKSLYVSATVILHSGSDMVQAERSGIPVTSFYQHFTKT 360  
DB 301 VLSRKVLDDGQNPRAEDLVGKSLYVSATVILHSGSDMVQAERSGIPVTSFYQHFTKT 360  
QY 361 PKYKPGMPPDLVFTNPDGSPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSQKPL 420  
DB 361 PKYKPGMPPDLVFTNPDGSPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSQKPL 420  
QY 421 SITVTRKQELSEAEQATRTMQALPYSTGVNSNNYHLVSLRTELPGFETLVNFFLRMD 480  
DB 421 SITVTRKQELSEAEQATRTMQALPYSTGVNSNNYHLVSLRTELPGFETLVNFFLRMD 480  
QY 481 RAHEAKRYTYLTMNKRLLKAGQVREPGQDLVLPISITITDFIPSPRLVAYTYLGA 540  
DB 481 RAHEAKRYTYLTMNKRLLKAGQVREPGQDLVLPISITITDFIPSPRLVAYTYLGA 540  
QY 541 SGQREVADSVWVDKSCVGSLLVYKSGQSEDRQPVPGQMTLKIEGDHGARVWLVAVDK 600  
DB 541 SGQREVADSVWVDKSCVGSLLVYKSGQSEDRQPVPGQMTLKIEGDHGARVWLVAVDK 600  
QY 601 GVFLVKNKLTOSKIDWVEKADIGCTPGSGKVAGVFSADGLTFTSSGQQTQRAEL 660  
DB 601 GVFLVKNKLTOSKIDWVEKADIGCTPGSGKVAGVFSADGLTFTSSGQQTQRAEL 660  
QY 661 QCPQFAARRRSVOLTETKMDKVGKYPKELRKCCEDGMRENPMRFSCQRTTFISLGEAC 720  
DB 661 QCPQFAARRRSVOLTETKMDKVGKYPKELRKCCEDGMRENPMRFSCQRTTFISLGEAC 720

QY 721 KKVFLDCCNYITELRQHRASHLGLARSNLDEDIAENIVSRSEFPESMLWNVEDLKE 780  
DB 721 KKVFLDCCNYITELRQHRASHLGLARSNLDEDIAENIVSRSEFPESMLWNVEDLKE 780  
QY 781 PPKNGISTKLNNIFLKDSTITTWEILAVSMDSKKGICVADPPFVTVMQDFFIDLRPLYSVV 840  
DB 781 PPKNGISTKLNNIFLKDSTITTWEILAVSMDSKKGICVADPPFVTVMQDFFIDLRPLYSVV 840  
QY 841 RNQVEIRAVLYNQRONQELKVRVELLHNPACSLATTKRHOQITITPPKSSLSVPVVI 900  
DB 841 RNQVEIRAVLYNQRONQELKVRVELLHNPACSLATTKRHOQITITPPKSSLSVPVVI 900  
QY 901 VPLKTLQGEVEKAAVYHHFISDGVKRSKLVVPEGRMNKTVAVETLDPERLREGVQKE 960  
DB 901 VPLKTLQGEVEKAAVYHHFISDGVKRSKLVVPEGRMNKTVAVETLDPERLREGVQKE 960  
QY 961 DIPADLSQDVPDTESEFILLQGTVAQMTEDAVDAERLKHLLVTPSGCGEONMIGTTP 1020  
DB 961 DIPADLSQDVPDTESEFILLQGTVAQMTEDAVDAERLKHLLVTPSGCGEONMIGTTP 1020  
QY 1021 TVIAVHYLDBTEQWEKFGLEKQGALELIKKGYYTQQLAFROPSSAFAAFVRAPESTWLTA 1080  
DB 1021 TVIAVHYLDBTEQWEKFGLEKQGALELIKKGYYTQQLAFROPSSAFAAFVRAPESTWLTA 1080  
QY 1081 YVVKVPSLAVNLIAIDSVLCGAVKWLILKOKPGVQEDAPVTHOEMI GGLRNNKED 1140  
DB 1081 YVVKVPSLAVNLIAIDSVLCGAVKWLILKOKPGVQEDAPVTHOEMI GGLRNNKED 1140  
QY 1141 MALTAFLVLSQBAKDCIEEQVNSLPGSITKAGDLEANYMNLQRSYTVAIAGYALAQMG 1200  
DB 1141 MALTAFLVLSQBAKDCIEEQVNSLPGSITKAGDLEANYMNLQRSYTVAIAGYALAQMG 1200  
QY 1201 RLKGPLLNKFLTTAKDKNWEDEPGKLVNVEATSVALLALQLQDQDFVPPVVRWLNQOR 1260  
DB 1201 RLKGPLLNKFLTTAKDKNWEDEPGKLVNVEATSVALLALQLQDQDFVPPVVRWLNQOR 1260  
QY 1261 YVGGYGSTOATPMVFOALAQYQKADPHOELNLDVSLQPSRSSKITHRIHWEASALLR 1320  
DB 1261 YVGGYGSTOATPMVFOALAQYQKADPHOELNLDVSLQPSRSSKITHRIHWEASALLR 1320  
QY 1321 SEETKNEGTVAEGKGGGTLSVVTMTHAKADQLTCKNPKLKYIKPAPETEKRPQDA 1380  
DB 1321 SEETKNEGTVAEGKGGGTLSVVTMTHAKADQLTCKNPKLKYIKPAPETEKRPQDA 1380  
QY 1381 KNTMILEICTRYRGDQDATMSILDISMMTGFAPDQDLDKQLANGVDYRISKYELDKAFSD 1440  
DB 1381 KNTMILEICTRYRGDQDATMSILDISMMTGFAPDQDLDKQLANGVDYRISKYELDKAFSD 1440  
QY 1441 RNTLIYLDKVSHEDDCLAFKVHOFVVELLTOPGAVKYAYNLEESCTRFYHPEKEDG 1500  
DB 1441 RNTLIYLDKVSHEDDCLAFKVHOFVVELLTOPGAVKYAYNLEESCTRFYHPEKEDG 1500  
QY 1501 KLNKLCRDELRCRCAEENCFIQSDDKVTLEERLDRKACBPGVDYVYKTLVQLSNDPDE 1560  
DB 1501 KLNKLCRDELRCRCAEENCFIQSDDKVTLEERLDRKACBPGVDYVYKTLVQLSNDPDE 1560  
QY 1561 YIMAIETIKSGSDVYVQCGQRTFTSPIKCREALKEBKHYLMGLSDPWEKPNLSY 1620  
DB 1561 YIMAIETIKSGSDVYVQCGQRTFTSPIKCREALKEBKHYLMGLSDPWEKPNLSY 1620  
QY 1621 IIGKDTWHEHPDEECQDEENKQKQCDLGAFTESMVVFGCPN 1663  
DB 1621 IIGKDTWHEHPDEECQDEENKQKQCDLGAFTESMVVFGCPN 1663

RESULT 6

AAW34618

ID AAW34618 standard; protein; 1663 AA.

XX AC AAW34618;

XX XX

DT 09-APR-1998 (first entry)



Db 1201 RLKGLLNKFLTTAKDKNRHEDPKQLYNYEATSYALLALLQLKDFDPVPPVVRWLNBOE 1260  
 Qy 1261 YGGGYSSTQATFWFVQALAQYQKADPDHQLNDVSLQPSRSSKITHRIHWESASLLR 1320  
 Db 1261 YGGGYSSTQATFWFVQALAQYQKADPDHQLNDVSLQPSRSSKITHRIHWESASLLR 1320  
 Qy 1321 SEETKENEFTVTAEGKGQGTLSVVTWYHAKAKDQLTCKNFDLKVTIKPAPETEKRPQDA 1380  
 Db 1321 SEETKENEFTVTAEGKGQGTLSVVTWYHAKAKDQLTCKNFDLKVTIKPAPETEKRPQDA 1380  
 Qy 1381 KNTWILEICRYRQDQATWSIIDISMTGFAPDQDQLQANGVDVRYISKYELDKAFSD 1440  
 Db 1381 KNTWILEICRYRQDQATWSIIDISMTGFAPDQDQLQANGVDVRYISKYELDKAFSD 1440  
 Qy 1441 RNTLIIVLDKVSHEDDCLAFKHQYFNVELIQPGAVKYVAYNLEESCTRFVHPEKEDG 1500  
 Db 1441 RNTLIIVLDKVSHEDDCLAFKHQYFNVELIQPGAVKYVAYNLEESCTRFVHPEKEDG 1500  
 Qy 1501 KLNKLCRDELRCRAENCFIQKSDDKVTLBERLDKACFQGVYVYKTRLVKQLSNDPDE 1560  
 Db 1501 KLNKLCRDELRCRAENCFIQKSDDKVTLBERLDKACFQGVYVYKTRLVKQLSNDPDE 1560  
 Qy 1561 YMAIEQTIKSGSDEVQVGQRTFISPIKCREALKLEEKHYLMWGLSSDFWGEKPNLSY 1620  
 Db 1561 YMAIEQTIKSGSDEVQVGQRTFISPIKCREALKLEEKHYLMWGLSSDFWGEKPNLSY 1620  
 Qy 1621 IIGKDTWVHWPEDECODEENKQCODLGAFTESMVWFGCPN 1663  
 Db 1621 IIGKDTWVHWPEDECODEENKQCODLGAFTESMVWFGCPN 1663

RESULT 7  
 AAM40989  
 ID AAM40989 standard; protein; 1663 AA.  
 AC AAM40989;  
 XX  
 DT 09-APR-1998 (first entry)  
 DE Human C3 protein mutant R1303X.  
 XX  
 KW Human; C3 protein; convertase; complement pathway protein; infection;  
 KW down-regulation resistant C3 convertase; xenograft rejection; therapy;  
 KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour;  
 KW complement-mediated response; MHC-mismatched lymphocyte; mutein.  
 XX  
 OS Homo sapiens.  
 XX  
 Key Location/Qualifiers  
 FH Misc-difference 1303  
 ET /label= Glu, Gln, Gly  
 XX  
 PN WO9732981-A1.  
 XX  
 PD 12-SEP-1997.  
 XX  
 PF 04-MAR-1997; 97WO-GB000603.  
 XX  
 PR 07-MAR-1996; 96GB-00004865.  
 PR 07-JUN-1996; 96GB-00011896.  
 PR 08-JUL-1996; 96GB-00014293.  
 PR 19-NOV-1996; 96GB-00024028.  
 XX  
 PA (IMUT-) IMUTRAN LTD.  
 XX  
 PI Farries TC, Harrison RA;  
 XX  
 DR WPI; 1997-457534/42.  
 XX  
 PT Modified complement pathway protein that forms C3 convertase resistant to  
 PT down-regulation - used to exhaust the complement pathway by super-  
 PT activation, especially for preventing graft rejection, etc.  
 XX

PS Claim 8; Page; 123pp; English.

XX This sequence represents a mutated human C3 protein of the invention (see  
 CC AAM34606 for wild type protein). This protein is a protein of the  
 CC invention, and is a modified native complement pathway protein (A) that  
 CC forms a down-regulation resistant C3 convertase. (A), their variants,  
 CC fragments and conjugates are used to deplete levels of complement pathway  
 CC proteins (by superactivation until one or more components are exhausted),  
 CC specifically to prevent rejection of foreign material (particularly a  
 CC xenograft) but also to prevent complement-mediated diseases resulting  
 CC from (surgical) injury or antibody-antigen interaction in autoimmune  
 CC disease, also to localise and/or amplify endogenous complement protein  
 CC conversion and deposition at a specific site (e.g. a virus, infected cell  
 CC or tumour, to increase sensitivity to complement-mediated responses; a  
 CC surgical removal of a tumour). Also contemplated is ex vivo treatment,  
 CC especially by passing blood through a matrix containing (A) (this may  
 CC remove additional anaphylactic peptides and other inflammatory mediators)  
 CC or killing of leukaemia cells or MHC-mismatched lymphocytes in extracted  
 CC bone marrow. Since (A) is not inhibited by factor I, it can bind  
 CC repeatedly to factor B (which is then inactivated), causing inactivation  
 CC of the alternative pathway by consumption of factor B

XX Sequence 1663 AA;

Query Match 99.9%; Score 8603; DB 2; Length 1663;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1662; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGPTSGPSLLLLLTHLPALGSPMYSIITPNILRSEETWVLEAHDAQGDVPVTVVH 60  
 Db 1 MGPTSGPSLLLLLTHLPALGSPMYSIITPNILRSEETWVLEAHDAQGDVPVTVVH 60  
 Qy 61 DFFGKKLVLSSEKTVLPATNMGVNTFTIPANREFKSEKGNKFPVTQATFGQVVEKV 120  
 Db 61 DFFGKKLVLSSEKTVLPATNMGVNTFTIPANREFKSEKGNKFPVTQATFGQVVEKV 120  
 Qy 121 VLVSQSGVLFQTDKTIYTPGSTVLYRIFTVNHKLLPVGRVTVMVNIENPEGIPVKQDSL 180  
 Db 121 VLVSQSGVLFQTDKTIYTPGSTVLYRIFTVNHKLLPVGRVTVMVNIENPEGIPVKQDSL 180  
 Qy 181 SSONQLGVPLSWDIPELVNMQWIKIRAYENSPQVSTFEFVKEYVLPSEFVIVEPTE 240  
 Db 181 SSONQLGVPLSWDIPELVNMQWIKIRAYENSPQVSTFEFVKEYVLPSEFVIVEPTE 240  
 Qy 241 KFYIYNEXGLEVTITARELYGKVEGTAFVIFGIDGQEQRISLPESLKRIPIEDGSGEV 300  
 Db 241 KFYIYNEXGLEVTITARELYGKVEGTAFVIFGIDGQEQRISLPESLKRIPIEDGSGEV 300  
 Qy 301 VLSRKVLDDGVONPRAEDLVGKSLVYSATVILHSGSDMWQAEBSGIPITVSPYQIHFTKT 360  
 Db 301 VLSRKVLDDGVONPRAEDLVGKSLVYSATVILHSGSDMWQAEBSGIPITVSPYQIHFTKT 360  
 Qy 361 PKYFKPGMPFDLMVFVTPNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHPSQKPL 420  
 Db 361 PKYFKPGMPFDLMVFVTPNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHPSQKPL 420  
 Qy 421 SITVTRTKQELSEAEQATRTMQALPYSTVGNSSNNYHLVRLTELRLPGETLNVNPLRMD 480  
 Db 421 SITVTRTKQELSEAEQATRTMQALPYSTVGNSSNNYHLVRLTELRLPGETLNVNPLRMD 480  
 Qy 481 RAHEAKIRYTYTILMNKGRLLKAGQVREPQGLVPLSLITTFPIPSFRVAVYTTIGA 540  
 Db 481 RAHEAKIRYTYTILMNKGRLLKAGQVREPQGLVPLSLITTFPIPSFRVAVYTTIGA 540  
 Qy 541 SGQREVVDVSVVDVKDSCVGLSVKSGQSDRQRPQGMQLKIEGDHGRVAVLVAVDK 600  
 Db 541 SGQREVVDVSVVDVKDSCVGLSVKSGQSDRQRPQGMQLKIEGDHGRVAVLVAVDK 600  
 Qy 601 GFVFLNKKNLKQSKINDVVERKADICCTPGSGKDYAGVFSAGLTFTSSSSQQAQRAEL 660  
 Db 601 GFVFLNKKNLKQSKINDVVERKADICCTPGSGKDYAGVFSAGLTFTSSSSQQAQRAEL 660

QY 661 QCQPAARRRRSQTLEKMDKVGKYPKELRKKCCEDGMRENPMRFSQQRTRFISLGEAC 720  
 DB 661 QCQPAARRRRSQTLEKMDKVGKYPKELRKKCCEDGMRENPMRFSQQRTRFISLGEAC 720  
 QY 721 KVFLOCCNYITELRQARASHGLARNSLDEDIIEENIVSRSEPEPEWLNVEDLKE 780  
 DB 721 KVFLOCCNYITELRQARASHGLARNSLDEDIIEENIVSRSEPEPEWLNVEDLKE 780  
 QY 781 PPKNGISTKLMNIFKOSITTWELAVNSDKKICVADPEVTVMDQDFIDLPLPSVV 840  
 DB 781 PPKNGISTKLMNIFKOSITTWELAVNSDKKICVADPEVTVMDQDFIDLPLPSVV 840  
 QY 841 RNEQVEIRAVLYNRYONQELKYRVLLHNPAFCSLATTKRRHQOTIIPPKSSLSVPYVI 900  
 DB 841 RNEQVEIRAVLYNRYONQELKYRVLLHNPAFCSLATTKRRHQOTIIPPKSSLSVPYVI 900  
 QY 901 VPLKTLQVEVEKAAVYHHFISDGVKSLKVVPEGIRNKNKTAVRTLDPERLGRGVQKE 960  
 DB 901 VPLKTLQVEVEKAAVYHHFISDGVKSLKVVPEGIRNKNKTAVRTLDPERLGRGVQKE 960  
 QY 961 DIPPADLSQVPTSETRIILQGTVPVQMTEDAVDAERLKHLLIVTPSGCGEQNMIGWTP 1020  
 DB 961 DIPPADLSQVPTSETRIILQGTVPVQMTEDAVDAERLKHLLIVTPSGCGEQNMIGWTP 1020  
 QY 1021 TVIAVHYLDETEQWEXFLEKKGALILEIKGYTQQLAFROPSSAPAAFKVRAPSTWLTA 1080  
 DB 1021 TVIAVHYLDETEQWEXFLEKKGALILEIKGYTQQLAFROPSSAPAAFKVRAPSTWLTA 1080  
 QY 1081 YVVKVFLANLIAIDSOVLCAVAKWLLILEKQKPGVFCEDAPVHQEMIGLGNNEKD 1140  
 DB 1081 YVVKVFLANLIAIDSOVLCAVAKWLLILEKQKPGVFCEDAPVHQEMIGLGNNEKD 1140  
 QY 1141 MALTAFLVLSQPAKIDCEQVNSLPGSITKAGDFLEANYNMQRSYTVAGVALAQMG 1200  
 DB 1141 MALTAFLVLSQPAKIDCEQVNSLPGSITKAGDFLEANYNMQRSYTVAGVALAQMG 1200  
 QY 1201 RLKGPLNLKFLTTAKDKNWEPPGKOLYNVEATSVALLALLQLKDFDVPVPEVWLNQOR 1260  
 DB 1201 RLKGPLNLKFLTTAKDKNWEPPGKOLYNVEATSVALLALLQLKDFDVPVPEVWLNQOR 1260  
 QY 1261 YCGGVGSGTOATPMVFPALAAQYKQADPHQELNLDVSLQLPSRSKITHRIHWSASLLR 1320  
 DB 1261 YCGGVGSGTOATPMVFPALAAQYKQADPHQELNLDVSLQLPSRSKITHRIHWSASLLR 1320  
 QY 1321 SEETKENEFTVTAEGKGQGTLSVVTMYHAKAKDQJTCNKDCLKVTIKPAPETEKRPQDA 1380  
 DB 1321 SEETKENEFTVTAEGKGQGTLSVVTMYHAKAKDQJTCNKDCLKVTIKPAPETEKRPQDA 1380  
 QY 1381 KNTMILEICTRYRGDQDATMSILDTSMTGFPAPDITDQLKLANGVDVRYISKYELDKAFSD 1440  
 DB 1381 KNTMILEICTRYRGDQDATMSILDTSMTGFPAPDITDQLKLANGVDVRYISKYELDKAFSD 1440  
 QY 1441 RNTLIYLDKVSSEDDCLAFKHQYFNVFVVELTQPGAVKYAYNLEESCTFRYPHEKEDG 1500  
 DB 1441 RNTLIYLDKVSSEDDCLAFKHQYFNVFVVELTQPGAVKYAYNLEESCTFRYPHEKEDG 1500  
 QY 1501 KLNKLRDELRCRAENCFIQSDDKVTLEBLDKACBPQVDVYVTKLVQVLSNDFDE 1560  
 DB 1501 KLNKLRDELRCRAENCFIQSDDKVTLEBLDKACBPQVDVYVTKLVQVLSNDFDE 1560  
 QY 1561 YMAIEQTIKSGSDEYVQVQGTFTSPIKCRALKLEEKKHVLMWGLSSDFWGEKPNLSY 1620  
 DB 1561 YMAIEQTIKSGSDEYVQVQGTFTSPIKCRALKLEEKKHVLMWGLSSDFWGEKPNLSY 1620  
 QY 1621 IIGKDTWEHWPBDEDCQENKQKQODLGAFTESWVFGCPN 1663  
 DB 1621 IIGKDTWEHWPBDEDCQENKQKQODLGAFTESWVFGCPN 1663

RESULT 8

AAW34608

ID AAW34608 standard; protein; 1663 AA.

XX

AC AAW34608;  
 XX 09-APR-1998 (first entry)  
 XX Human C3 protein mutant C3M-26.  
 XX Human; C3 protein; convertase; complement pathway protein; infection;  
 XX down-regulation resistant C3 convertase; xenograft rejection; therapy;  
 XX complement-mediated disease; autoimmune disease; leukaemia cell; tumour;  
 XX complement-mediated response; MHC-mismatched lymphocyte; muten.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 XX Misc-difference 1303 /note= "Arg to Gln mutation"  
 XX Misc-difference 1320 /note= "Arg to Gln mutation"  
 XX WO9732981-A1.  
 XX 12-SEP-1997.  
 XX 04-MAR-1997; 97WO-GB0006503.  
 XX 07-MAR-1996; 96GB-00004865.  
 XX 07-JUN-1996; 96GB-00011896.  
 XX 08-JUL-1996; 96GB-00014293.  
 XX 19-NOV-1996; 96GB-00024028.  
 XX (IMUT-) IMUTRAN LTD.  
 XX Farries TC, Harrison RA;  
 XX WPI; 1997-457534/42.  
 XX Modified complement pathway protein that forms C3 convertase resistant to  
 XX down-regulation - used to exhaust the complement pathway by super-  
 XX activation, especially for preventing graft rejection, etc.  
 XX Example 4; Page; 123pp; English.  
 XX This sequence represents a mutated human C3 protein of the invention (see  
 XX AAW34608 for wild type protein). This protein is a protein of the  
 XX invention, and is a modified native complement pathway protein (A) that  
 XX forms a down-regulation resistant C3 convertase. (A), their variants,  
 XX fragments and conjugates are used to deplete levels of complement pathway  
 XX proteins (by superactivation until one or more components are exhausted),  
 XX specifically to prevent rejection of foreign material (particularly a  
 XX xenograft) but also to prevent complement-mediated diseases resulting  
 XX from (surgical) injury or antibody-antigen interaction in autoimmune  
 XX disease, also to localise and/or amplify endogenous complement protein  
 XX conversion and deposition at a specific site (e.g. a virus, infected cell  
 XX or tumour, to increase sensitivity to complement-mediated responses; a  
 XX particular application is eliminating any cancer cells left after  
 XX surgical removal of a tumour). Also contemplated is ex vivo treatment,  
 XX especially by passing blood through a matrix containing (A) (this may  
 XX remove additional anaphylactic peptides and other inflammatory mediators)  
 XX or killing of leukaemia cells or MHC-mismatched lymphocytes in extracted  
 XX bone marrow. Since (A) is not inhibited by factor I, it can bind  
 XX repeatedly to factor B (which is then inactivated), causing inactivation  
 XX of the alternative pathway by consumption of factor B  
 XX Sequence 1663 AA;  
 XX Query Match 99.9%; Score 8601; DB 2; Length 1663;  
 XX Best Local Similarity 99.9%; Pred. No. 0;  
 XX Matches 1661; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPTSGPSLLLLLTHLPALGSPMYSIITPNILRLESEETWVLEAHDAGQDVPVTVVH 60  
 DB 1 MGPTSGPSLLLLLTHLPALGSPMYSIITPNILRLESEETWVLEAHDAGQDVPVTVVH 60

QY 61 DPGKKVLSSSEKTLTPATNMGNVTFTIPANREKSKGRNKFVTVQATFGTOVSEKV 120  
DB 61 DPGKKVLSSSEKTLTPATNMGNVTFTIPANREKSKGRNKFVTVQATFGTOVSEKV 120  
QY 121 VLVSQSGLPIQTDKTIYTPGSTVLVRIETVNHKLLPVGRITVMNIENPEGIPVKQDSL 180  
DB 121 VLVSQSGLPIQTDKTIYTPGSTVLVRIETVNHKLLPVGRITVMNIENPEGIPVKQDSL 180  
QY 181 SSQNLGVLPLSWDIPELVNMGWKIRAYYENSPQOVFTETPEVEKVEYVLPSEVIVEPTE 240  
DB 181 SSQNLGVLPLSWDIPELVNMGWKIRAYYENSPQOVFTETPEVEKVEYVLPSEVIVEPTE 240  
QY 241 KFYIYNEKGLVETITAFILYKKGVEGTAFVIFGIDGQRIISLPSLKRIPIEDGSGEV 300  
DB 241 KFYIYNEKGLVETITAFILYKKGVEGTAFVIFGIDGQRIISLPSLKRIPIEDGSGEV 300  
QY 301 VLSRKVLDDGVONPRAEDLVGKSLVYSATVILHSGSDMWQAERSGIPVITSFYQHFTKT 360  
DB 301 VLSRKVLDDGVONPRAEDLVGKSLVYSATVILHSGSDMWQAERSGIPVITSFYQHFTKT 360  
QY 361 PKYFKPGNPDLMVFTNPDGSPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSOKPL 420  
DB 361 PKYFKPGNPDLMVFTNPDGSPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSOKPL 420  
QY 421 SITVTRTKQELSEAEQATRTMOALFYSTVGNNSNYLHLSVLRTLPFGTLLNVNELLRMD 480  
DB 421 SITVTRTKQELSEAEQATRTMOALFYSTVGNNSNYLHLSVLRTLPFGTLLNVNELLRMD 480  
QY 481 RAHAKEIRYTYLIMNKGRLKAGQVREPQGLVPLSITTDRTPSRLVAVYTLGA 540  
DB 481 RAHAKEIRYTYLIMNKGRLKAGQVREPQGLVPLSITTDRTPSRLVAVYTLGA 540  
QY 541 SQREWVADSVVMDVKSCVGLSVKSGQSDRQPVPGQMTLKIEGDHGAARVILVAVDK 600  
DB 541 SQREWVADSVVMDVKSCVGLSVKSGQSDRQPVPGQMTLKIEGDHGAARVILVAVDK 600  
QY 601 GFVLNKKVLTQSKIWDVVEKADIGCTPGSGDYAGVFSDAQLTPTSSSGQQTQRAEL 660  
DB 601 GFVLNKKVLTQSKIWDVVEKADIGCTPGSGDYAGVFSDAQLTPTSSSGQQTQRAEL 660  
QY 661 QCQPAARRRSVQLTEKMDKVGPKELRKCECDGMRNPMRFSCQRTFISLGEAC 720  
DB 661 QCQPAARRRSVQLTEKMDKVGPKELRKCECDGMRNPMRFSCQRTFISLGEAC 720  
QY 721 KKVFLDCCNYITELRRQARASHLGLARSLNLDIEDIAENIVSRSEFPESLWNVEDLKE 780  
DB 721 KKVFLDCCNYITELRRQARASHLGLARSLNLDIEDIAENIVSRSEFPESLWNVEDLKE 780  
QY 781 PPQGISIKLNMIFLKOSITTWELAVMSDKKGI CVADPFVTVMQDFIDRLPYSVV 840  
DB 781 PPQGISIKLNMIFLKOSITTWELAVMSDKKGI CVADPFVTVMQDFIDRLPYSVV 840  
QY 841 RNEQVEIRAVLYNQELKVRVELLHNPFAFCSLATTKRHOQITTIIPKSSLSVPVI 900  
DB 841 RNEQVEIRAVLYNQELKVRVELLHNPFAFCSLATTKRHOQITTIIPKSSLSVPVI 900  
QY 901 VBLKTLGSEVKAAYVHHFISDGVRSKLVWPEGIRMNKTAVRDLPERLGRGVQKE 960  
DB 901 VBLKTLGSEVKAAYVHHFISDGVRSKLVWPEGIRMNKTAVRDLPERLGRGVQKE 960  
QY 961 DIPPADLSQVDPDTESETRILLQGTTPVAQMTEDAVDAERLKHILVTPSGCGQNMGTP 1020  
DB 961 DIPPADLSQVDPDTESETRILLQGTTPVAQMTEDAVDAERLKHILVTPSGCGQNMGTP 1020  
QY 1021 TVIANVHLEDETEQNEKGLKBOGALBLIKVGYTOQLAPQSSAAFAVKAPSTWTLTA 1080  
DB 1021 TVIANVHLEDETEQNEKGLKBOGALBLIKVGYTOQLAPQSSAAFAVKAPSTWTLTA 1080  
QY 1081 YVVKVFLAVNLIAIDSVLCGAVKWLILEKQKPDGVQFQEDAPVHCEMIGGLRNNKED 1140  
DB 1081 YVVKVFLAVNLIAIDSVLCGAVKWLILEKQKPDGVQFQEDAPVHCEMIGGLRNNKED 1140  
QY 1141 MALTAFLVLSIQEAKDICEEQVNSLPGSITKAGDFLEANNYMLQORSYTVAIAGYALQMG 1200

DB 1141 MALTAFLVLSIQEAKDICEEQVNSLPGSITKAGDFLEANNYMLQORSYTVAIAGYALQMG 1200  
QY 1201 RLKGPLLNKFLTTAKDKRWEDPGKQLYNVEATSYALLALLQLKDFDPVPPVRLNEQR 1260  
DB 1201 RLKGPLLNKFLTTAKDKRWEDPGKQLYNVEATSYALLALLQLKDFDPVPPVRLNEQR 1260  
QY 1261 YYGCGYGSTQATFMVFOALAQYQKADPDHQLNLDVSLQPLPSRSSKITHRIHWESASLLR 1320  
DB 1261 YYGCGYGSTQATFMVFOALAQYQKADPDHQLNLDVSLQPLPSRSSKITHRIHWESASLLR 1320  
QY 1321 SEETENEGFTVTAEGKGGTSLSVTVMYHAKAQDLTCNKFDLKVTIKPAPETEKPODA 1380  
DB 1321 SEETENEGFTVTAEGKGGTSLSVTVMYHAKAQDLTCNKFDLKVTIKPAPETEKPODA 1380  
QY 1381 KNTMLEICTRYRGQDATMSILDISMMTGFPAPDLDLKLQANGVDRYISKYELDKAFSD 1440  
DB 1381 KNTMLEICTRYRGQDATMSILDISMMTGFPAPDLDLKLQANGVDRYISKYELDKAFSD 1440  
QY 1441 RNTLIYLDKVSHEDDCLAFKHQYFNVELIQFQAVKYAYYNLBSCTRFVHPEKEDG 1500  
DB 1441 RNTLIYLDKVSHEDDCLAFKHQYFNVELIQFQAVKYAYYNLBSCTRFVHPEKEDG 1500  
QY 1501 KLNKLCRDELCLCAEENCFIQKSDDKVTLBERLDKACEPGVDVYVYKTRLVKQVLSNDFDE 1560  
DB 1501 KLNKLCRDELCLCAEENCFIQKSDDKVTLBERLDKACEPGVDVYVYKTRLVKQVLSNDFDE 1560  
QY 1561 YVMAIEQITKSGSDVQVQQTFTFISPKCREALKLEKHYLMWGLSSDFWGEKPNLSY 1620  
DB 1561 YVMAIEQITKSGSDVQVQQTFTFISPKCREALKLEKHYLMWGLSSDFWGEKPNLSY 1620  
QY 1621 IIGKOTVHWPHEDECOENQKOCQDLGAFTESMVVFQCPN 1663  
DB 1621 IIGKOTVHWPHEDECOENQKOCQDLGAFTESMVVFQCPN 1663

RESULT 9  
AAR94028  
ID AAR94028 standard; protein; 1663 AA.  
XX AAR94028;  
AC AAR94028;  
DT 21-MAY-1996 (first entry)  
XX Human C3 precursor.  
DE C3 protein; convertase; Factor I; Factor H; complement.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH Peptide 1..22  
FT /label= Sig\_peptide  
FT 23..667  
FT /note= "C3 beta chain"  
FT 668..671  
FT /note= "amino acids 668-671 are removed when the  
FT precursor is cleaved into the alpha and beta chains"  
FT 672..1663  
FT /note= "C3 alpha chain"  
XX WO9607738-A2.  
XX 14-MAR-1996.  
XX 08-SEP-1995; 95WO-CB002121.  
XX 08-SEP-1994; 94GB-00018147.  
XX 04-MAY-1995; 95GB-00009102.  
XX (IMUT-) IMUTRAN LTD.  
XX Harrison RA, Farries TC;  
PI

XX WPI; 1996-171613/17.  
 DR N-PSDB; AAT17738.  
 XX  
 PT Mutant complement pathway protein forming stable C3 convertase - for  
 PT Generalised complement depletion or localised complement activation.  
 XX  
 PS Disclosure; Fig 1; 81pp; English.  
 XX  
 CC Human C3 protein (AAR94028) was produced by expression of a cDNA sequence  
 CC (AAT17738) isolated from a human liver cDNA library. C3 is a complement  
 CC pathway protein that is susceptible to cleavage by Factor I and is also  
 CC susceptible to the inhibitory action of Factor H. Mutants of C3 (AAR94029  
 CC and AAR94030) have been produced by site-directed mutagenesis. These  
 CC mutants can be used to super-activate the complement system, or to induce  
 CC localised super-activation at a specific target to increase the target's  
 CC sensitivity to complement-mediated destruction  
 XX  
 XX Sequence 1663 AA;  
 SQ

Query Match 99.9%; Score 8600; DB 2; Length 1663;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1662; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGPTSGPSLLLLTHPLALGSPMYSIIITPNILRESEETWVLEAHDAQDQVPTVTVH 60  
 DB 1 MGPTSGPSLLLLTHPLALGSPMYSIIITPNILRESEETWVLEAHDAQDQVPTVTVH 60

QY 61 DFPKKVLVSEKTVITPATNHGNTFTIPANRSEKSEKGNKFPVTVQATFGQVVEKV 120  
 DB 61 DFPKKVLVSEKTVITPATNHGNTFTIPANRSEKSEKGNKFPVTVQATFGQVVEKV 120

QY 121 VLVSQSGYLFIQDTKTIYTPGSTVLYRIFTVNHKLLPVGRVMVNIENPGIIPVKQDSL 180  
 DB 121 VLVSQSGYLFIQDTKTIYTPGSTVLYRIFTVNHKLLPVGRVMVNIENPGIIPVKQDSL 180

QY 181 SSQNLGVPLSDWIDPELVNMGQWKIRAYENSPQCVFSTEBEVKBYVLPSEFVIVEPTE 240  
 DB 181 SSQNLGVPLSDWIDPELVNMGQWKIRAYENSPQCVFSTEBEVKBYVLPSEFVIVEPTE 240

QY 241 KFYIYNEKGLVITITARFLYKKGVEGTAFVIFGIDGQERISLPESLKRIPIEDGSGEV 300  
 DB 241 KFYIYNEKGLVITITARFLYKKGVEGTAFVIFGIDGQERISLPESLKRIPIEDGSGEV 300

QY 301 VLSRKVLVDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQAESGIPVTSPIQHFTKT 360  
 DB 301 VLSRKVLVDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQAESGIPVTSPIQHFTKT 360

QY 361 PKYFKPGMPDLVFTNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHPSQKPL 420  
 DB 361 PKYFKPGMPDLVFTNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHPSQKPL 420

QY 421 SITVTRTKKSELSAEQATRMQALPYSTVGNSENLYLHLSVLTRELPGSTLNNFLRMD 480  
 DB 421 SITVTRTKKSELSAEQATRMQALPYSTVGNSENLYLHLSVLTRELPGSTLNNFLRMD 480

QY 481 RAHEAKIRYTYILIMNGKRLLAGROVREPQGLVPLSLITTFIPSRFLVAYTYLIGA 540  
 DB 481 RAHEAKIRYTYILIMNGKRLLAGROVREPQGLVPLSLITTFIPSRFLVAYTYLIGA 540

QY 541 SGQREVVDSDVVDKSDCVGSLVKSQSGEDRQVPVQGMQLKIEGDHGARVVLVAVDK 600  
 DB 541 SGQREVVDSDVVDKSDCVGSLVKSQSGEDRQVPVQGMQLKIEGDHGARVVLVAVDK 600

QY 601 GVFLNKKKLTQSKTDWVVEKADIGCTPGSGOKVAGVPSDAGLTFPTSSSQQTQAQRAEL 660  
 DB 601 GVFLNKKKLTQSKTDWVVEKADIGCTPGSGOKVAGVPSDAGLTFPTSSSQQTQAQRAEL 660

QY 661 QCPOPAARRRRSQTLEKMDKVGKYPKELRKCCEGDMRENPMRPFSCQRTTRFISLGEAC 720  
 DB 661 QCPOPAARRRRSQTLEKMDKVGKYPKELRKCCEGDMRENPMRPFSCQRTTRFISLGEAC 720

QY 721 KKVFLDCCNITITELRQHARASHGLARSLNDEDAEENIVSRSEFFESLMWVEDLKE 780

DB 721 KKVFLDCCNITITELRQHARASHGLARSLNDEDAEENIVSRSEFFESLMWVEDLKE 780  
 QY 781 PPKNGISTKLMNIFLKDSTITTEILAVSMDKKGICVADPPPEVTYMQDFFIDLRLPSVY 840  
 DB 781 PPKNGISTKLMNIFLKDSTITTEILAVSMDKKGICVADPPPEVTYMQDFFIDLRLPSVY 840

QY 841 RNEQVEIRAVLYNVRQNOELKURVELLHNPAFCSLATTKRHQQTITTPKSSISVPVI 900  
 DB 841 RNEQVEIRAVLYNVRQNOELKURVELLHNPAFCSLATTKRHQQTITTPKSSISVPVI 900

QY 901 VPLTKGLQEVVEKAAVYHHFISDGVKSKLVVPEGIRMNKTVAVRTLPDLREGVQKE 960  
 DB 901 VPLTKGLQEVVEKAAVYHHFISDGVKSKLVVPEGIRMNKTVAVRTLPDLREGVQKE 960

QY 961 DIPADLSQDVPTESESTRILQGTTPVAQWTDVAERLKHILVTPSGCGEQNMIGWTP 1020  
 DB 961 DIPADLSQDVPTESESTRILQGTTPVAQWTDVAERLKHILVTPSGCGEQNMIGWTP 1020

QY 1021 TVIAVHYLDETEQWEKFGLEKQGALELIKGYTQOLAFPROPSAFAAFVRAESTWLT 1080  
 DB 1021 TVIAVHYLDETEQWEKFGLEKQGALELIKGYTQOLAFPROPSAFAAFVRAESTWLT 1080

QY 1081 YVWVFSLANVLLAIDQSQVLCGAVKWLILEKQKPGVQEDAPVTHQSMIGLNNNEKD 1140  
 DB 1081 YVWVFSLANVLLAIDQSQVLCGAVKWLILEKQKPGVQEDAPVTHQSMIGLNNNEKD 1140

QY 1141 MALTAFAVLIISLOBAKOICEBOVNSLPGSTITKAGDFLEANYMLQBSYVAIAGYALQMG 1200  
 DB 1141 MALTAFAVLIISLOBAKOICEBOVNSLPGSTITKAGDFLEANYMLQBSYVAIAGYALQMG 1200

QY 1201 RLKGPLLNKFLTTAKQKNRWEDPGKQLYNVEATSYALLALLQKDFDVPVFWLWNEQ 1260  
 DB 1201 RLKGPLLNKFLTTAKQKNRWEDPGKQLYNVEATSYALLALLQKDFDVPVFWLWNEQ 1260

QY 1261 YGCGYSTQATFMVFOALAQYKQDAPDHOELNDVLSQPSRSSKITRHHWSASLLR 1320  
 DB 1261 YGCGYSTQATFMVFOALAQYKQDAPDHOELNDVLSQPSRSSKITRHHWSASLLR 1320

QY 1321 SEETKENEGTVAEKGQGTLSVVTMTHAKAKDQLTCKNFDLKVTIKPAPETEKRPQDA 1380  
 DB 1321 SEETKENEGTVAEKGQGTLSVVTMTHAKAKDQLTCKNFDLKVTIKPAPETEKRPQDA 1380

QY 1381 KNTMILICTRYGDDQDATMSILDISMTGFPADTDDLKOLANGVDVYISKVELDKAFSD 1440  
 DB 1381 KNTMILICTRYGDDQDATMSILDISMTGFPADTDDLKOLANGVDVYISKVELDKAFSD 1440

QY 1441 RNTLIIYLDKVSHEDDCLAFKVHGFVNYVELIQGAVKVYAYYNLEESCTFYHPEKEDG 1500  
 DB 1441 RNTLIIYLDKVSHEDDCLAFKVHGFVNYVELIQGAVKVYAYYNLEESCTFYHPEKEDG 1500

QY 1501 KLNKLCRDELRCRCAEENCFIQKSDDDKVTLEERLDKACBPQVDVYKTLVKVQSLNDFE 1560  
 DB 1501 KLNKLCRDELRCRCAEENCFIQKSDDDKVTLEERLDKACBPQVDVYKTLVKVQSLNDFE 1560

QY 1561 YIMAIEQTIKSGSEVQVGQRTFISPIKCRBALKEKKHYLMWGLSDFWKGKPNLSY 1620  
 DB 1561 YIMAIEQTIKSGSEVQVGQRTFISPIKCRBALKEKKHYLMWGLSDFWKGKPNLSY 1620

QY 1621 IIGKDTVWEHWPBEDECOBENQKQODLGAFETSMVWFGCPN 1663  
 DB 1621 IIGKDTVWEHWPBEDECOBENQKQODLGAFETSMVWFGCPN 1663

RESULT 10  
 AAW34627  
 ID AAW34627 standard; protein; 1663 AA.  
 XX  
 AC AAW34627;  
 XX  
 DT 09-APR-1998 (first entry)  
 XX  
 DE Human C3 protein mutant FT-5.

XX Human; C3 protein; convertase; complement pathway protein; infection; 181 SSONQLGVPLSWDIPELVNMQWIKIRAYENSPQOVSTEFVKEVYVLPSEFVIVPEPTE 240  
 KW down-regulation resistant C3 convertase; xenograft rejection; therapy; 181 SSONQLGVPLSWDIPELVNMQWIKIRAYENSPQOVSTEFVKEVYVLPSEFVIVPEPTE 240  
 KW complement-mediated disease; autoimmune disease; leukaemia cell; tumour; 241 KFYIYINKEGLEVTITARFLYKGVKVEGTAFVIFGIDGEGEQRISLPESLKRIPIEDGSGEV 300  
 XX complement-mediated response; MHC-mismatched lymphocyte; mutein. 241 KFYIYINKEGLEVTITARFLYKGVKVEGTAFVIFGIDGEGEQRISLPESLKRIPIEDGSGEV 300  
 OS Homo sapiens.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1661 /note= "C1661S mutation"  
 XX  
 XX WO9732981-A1.  
 XX 12-SEP-1997.  
 XX  
 XX 04-MAR-1997; 97WO-GB000603.  
 XX  
 XX 07-MAR-1996; 96GB-00004865.  
 PR 07-JUN-1996; 96GB-00011896.  
 PR 08-JUL-1996; 96GB-00014293.  
 PR 19-NOV-1996; 96GB-00024028.  
 XX  
 XX (IMUT-) IMUTRAN LTD.  
 XX  
 XX Farries TC, Harrison RA;  
 PI WPI; 1997-457534/42.  
 DR  
 XX  
 XX Modified complement pathway protein that forms C3 convertase resistant to 601 GFVLNKNKLTQSKIWDVVERADIGCTPGSGKYAGVFSAGLTFTSSSQOQTAQRAEL 660  
 PT down-regulation - used to exhaust the complement pathway by super- 601 GFVLNKNKLTQSKIWDVVERADIGCTPGSGKYAGVFSAGLTFTSSSQOQTAQRAEL 660  
 PT activation, especially for preventing graft rejection, etc.  
 XX  
 XX Example 17; Page; 123pp; English.  
 PS  
 XX  
 XX This sequence represents a mutated human C3 protein of the invention (see 661 QCPQPAARRRSVOLTETKEMDKVGPKELEKCCEDGMRNEMBPSCORRTRFISLGEAC 720  
 CC AW34606 for wild type protein). This protein is a protein of the 661 QCPQPAARRRSVOLTETKEMDKVGPKELEKCCEDGMRNEMBPSCORRTRFISLGEAC 720  
 CC invention, and is a modified native complement pathway protein (A) that 721 KVFVLDCCNYITELRQRHARASHLGLARSLNDELIIAENIVSRSEPPESWLMVVDLKE 780  
 CC forms a down-regulation resistant C3 convertase. (A), their variants, 721 KVFVLDCCNYITELRQRHARASHLGLARSLNDELIIAENIVSRSEPPESWLMVVDLKE 780  
 CC fragments and conjugates are used to deplete levels of complement pathway 781 PPKNGSTLKMNIFLKDSITTWELLAVSMOKKGCIVADPEVTVMDOFFIDRLPYSVW 840  
 CC proteins (by superactivation until one or more components are exhausted), 781 PPKNGSTLKMNIFLKDSITTWELLAVSMOKKGCIVADPEVTVMDOFFIDRLPYSVW 840  
 CC specifically to prevent rejection of foreign material (particularly a 841 RNEQVEIRAVLYNRQNOELKVRVLLHNPAFCSLATTKRHHQQTITIPPXSSLSVPYVI 900  
 CC xenograft) but also to prevent complement-mediated diseases resulting 841 RNEQVEIRAVLYNRQNOELKVRVLLHNPAFCSLATTKRHHQQTITIPPXSSLSVPYVI 900  
 CC from (surgical) injury or antibody-antigen interaction in autoimmune 901 VPLKTLGLEVEVKAAVYHHFISDGVKSLKVVEGIRMNKTAVRTLDPERLGRGVQKE 960  
 CC diseases, also to localise and/or amplify endogenous complement protein 901 VPLKTLGLEVEVKAAVYHHFISDGVKSLKVVEGIRMNKTAVRTLDPERLGRGVQKE 960  
 CC conversion and deposition at a specific site (e.g. a virus, infected cell 961 DIPPADLSQVDPDTESETRILLQGTPTVAQMTEDAVDAERLKHILVTPSGCGEQNMIGWTP 1020  
 CC or tumour, to increase sensitivity to complement-mediated responses; a 961 DIPPADLSQVDPDTESETRILLQGTPTVAQMTEDAVDAERLKHILVTPSGCGEQNMIGWTP 1020  
 CC particular application is eliminating any cancer cells left after 1021 TVIAVHVLDETQWEKFGLEKEQGALELIKKGYTQQLAFROPSSAFARAPSTWLTA 1080  
 CC surgical removal of a tumour). Also contemplated is ex vivo treatment, 1021 TVIAVHVLDETQWEKFGLEKEQGALELIKKGYTQQLAFROPSSAFARAPSTWLTA 1080  
 CC especially by passing blood through a matrix containing (A) (this may, 1081 YVVKFSLAVNLIAIDSQVLCGAVKWLILEKQKPDGVFQEDAPVHQBEMIGGLRNNNEKD 1140  
 CC remove additional anaphylactic peptides and other inflammatory mediators) 1081 YVVKFSLAVNLIAIDSQVLCGAVKWLILEKQKPDGVFQEDAPVHQBEMIGGLRNNNEKD 1140  
 CC or killing of leukaemia cells or MHC-mismatched lymphocytes in extracted 1141 MALTAFLVLSLOEAKDICEEONSLPGSITKAGDFLEANYMNIORSYTVATAGYALQMG 1200  
 CC bone marrow. Since (A) is not inhibited by factor I, it can bind 1141 MALTAFLVLSLOEAKDICEEONSLPGSITKAGDFLEANYMNIORSYTVATAGYALQMG 1200  
 CC repeatedly to factor B (which is then inactivated), causing inactivation 1201 RLKGPLLNKFLTTAKDKRWEDPGKQLYNVEATSYALLALILQLKDFDPPVPPVRLNEQR 1260  
 CC of the alternative pathway by consumption of factor B 1201 RLKGPLLNKFLTTAKDKRWEDPGKQLYNVEATSYALLALILQLKDFDPPVPPVRLNEQR 1260  
 XX  
 XX Sequence 1663 AA;  
 SQ  
 Query Match 99.98; Score 8599; DB 2; Length 1663;  
 Best Local Similarity 99.98; Pred. No. 0;  
 Matches 1662; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MGPTSGPSLLLLTHPLALGSPMYSIITPNILRLESEETMVLRAHDAQGDPVTVTVH 60  
 DB 1 MGPTSGPSLLLLTHPLALGSPMYSIITPNILRLESEETMVLRAHDAQGDPVTVTVH 60  
 QY 61 DFGCKVLVLSSEKTLTPATNMGNTVFTIPANREKSEKGNKFTVCAFTGTVQVEKY 120  
 DB 61 DFGCKVLVLSSEKTLTPATNMGNTVFTIPANREKSEKGNKFTVCAFTGTVQVEKY 120  
 QY 121 VLVSQSGYLFQTDKTIYTPGTVLYRIFTVNHKLLPVGRVTVMNIENPEGIPVKQDSL 180  
 DB 121 VLVSQSGYLFQTDKTIYTPGTVLYRIFTVNHKLLPVGRVTVMNIENPEGIPVKQDSL 180



QY 1261 YGGGYSQTATFWMVFOALAQYQKADPDHQLNLDVSLQPSRSSKITHRIHWESASLLR 1320  
Db 1261 YGGGYSQTATFWMVFOALAQYQKADPDHQLNLDVSLQPSRSSKITHRIHWESASLLR 1320  
QY 1321 SEETKNEGTVTAEGKGQTLVWVTHYAKAKDQLTCKNPKLVTKIAPATEKRPQDA 1380  
Db 1321 SEETKNEGTVTAEGKGQTLVWVTHYAKAKDQLTCKNPKLVTKIAPATEKRPQDA 1380  
QY 1381 KNTMILBICTRYRGDQATMSIILDISMTGFAFPTDLDKQLANGVDVRIKVELDKAFSD 1440  
Db 1381 KNTMILBICTRYRGDQATMSIILDISMTGFAFPTDLDKQLANGVDVRIKVELDKAFSD 1440  
QY 1441 RNTLIIYLDKVSHEDDCLAFKHQVFNVELIQPGAKVYAYNLEESCTFYHPEKEDG 1500  
Db 1441 RNTLIIYLDKVSHEDDCLAFKHQVFNVELIQPGAKVYAYNLEESCTFYHPEKEDG 1500  
QY 1501 KLNLCDLRCRAEENCFIQKSDDKVTLLESLDRACRPGVDYVYKTLVKVQLSNDPDE 1560  
Db 1501 KLNLCDLRCRAEENCFIQKSDDKVTLLESLDRACRPGVDYVYKTLVKVQLSNDPDE 1560  
QY 1561 YIMAEOTIKSGDEVOVGQORTFISPIKREALKLEBKHYLWGLSSDFWGEKPNLSY 1620  
Db 1561 YIMAEOTIKSGDEVOVGQORTFISPIKREALKLEBKHYLWGLSSDFWGEKPNLSY 1620  
QY 1621 IIGKDTVWEHWPBEDECOBENKQCODLGAFTESMVVFGCPN 1663  
Db 1621 IIGKDTVWEHWPBEDECOBENKQCODLGAFTESMVVFGSPN 1663

RESULT 11

ADB90023

ID ADB90023 standard; protein; 1663 AA.

AC ADB90023;

DT 04-DEC-2003 (first entry)

XX House complement component C3.

DE Human; antisense; complement component C3; inflammation; septic shock;  
KW multiple organ failure; hyperacute organ failure; autoimmune disorder;  
KW CNS inflammation; multiple sclerosis; atherosclerosis; tumour.

XX Homo sapiens.

OS US2003096775-A1.

PN 22-MAY-2003.

PD 23-OCT-2001; 2001US-00001076.

PF 23-OCT-2001; 2001US-00001076.

PR (ISIS-) ISIS PHARM INC.

XX Graham MJ, Watt AT;

XX WPI; 2003-606441/57.

DR N-PSDB; ADB9847.

XX New antisense oligonucleotides targeted to a nucleic acid molecule  
PT encoding complement component C3, useful for treating a disease or  
PT condition associated with complement component C3, e.g. autoimmune  
PT disorder or infection.

XX Example 13; Page 28-34; 72pp; English.

CC The invention relates to a compound 8-50 nucleobases in length targeted  
CC to a nucleic acid molecule encoding complement component C3. The compound  
CC specifically hybridises with the nucleic acid molecule encoding  
CC complement component C3 and inhibits the expression of complement  
CC component C3, or specifically hybridises with at least an 8-nucleobase  
CC portion of an active site on a nucleic acid molecule encoding complement

CC component C3. Also included are a composition comprising the compound and  
CC a pharmaceutical carrier or diluent, inhibiting the expression of  
CC complement component C3 in cells or tissues (comprising contacting the  
CC cells or tissues with the compound cited above) and treating an animal  
CC having a disease or condition associated with complement component C3  
CC comprising administering to the animal the compound cited above so that  
CC expression of complement component C3 is inhibited. The antisense  
CC compounds are useful for inhibiting the expression of complement  
CC component C3 in cells or tissues, or for treating an animal having a  
CC disease or condition associated with complement component C3 such as an  
CC autoimmune disorder (e.g. multiple sclerosis), an infection or  
CC atherosclerosis, inflammation, septic shock, multiple organ failure,  
CC hyperacute organ failure and CNS inflammation. The compounds are also  
CC useful as research reagents and diagnostics, in distinguishing functions  
CC of various members of a biological pathway, or for preventing or delaying  
CC infection, inflammation or tumour formation. The present sequence is a  
CC complement component C3 protein sequence.

XX Sequence 1663 AA;

Query Match 22.98; Score 8598; DB 7; Length 1663;

Best Local Similarity 99.94; Pred. NO. 0;

Matches 1661; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGPTSGPSLLLLTHLPALGSPMYSIITPNILRESEETVLEBAHQDQDVPVTVVH 60

Db 1 MGPTSGPSLLLLTHLPALGSPMYSIITPNILRESEETVLEBAHQDQDVPVTVVH 60

QY 61 DPGKCLVLSSEKTVLPATNHNQVTFIPANREFKSEKGNKVTVOATFGTVQVEKV 120

Db 61 DPGKCLVLSSEKTVLPATNHNQVTFIPANREFKSEKGNKVTVOATFGTVQVEKV 120

QY 121 VLVSQSGYLFIQDQKTIYTPGSTVLYRIFTVNNHLLPVGRTVMVNIENPEGIPVKQDSL 180

Db 121 VLVSQSGYLFIQDQKTIYTPGSTVLYRIFTVNNHLLPVGRTVMVNIENPEGIPVKQDSL 180

QY 181 SSQNLGVLPISWDIPELVNMGWKIRAYENSPOQVSTEFVEKVEYLPFVIVEPTE 240

Db 181 SSQNLGVLPISWDIPELVNMGWKIRAYENSPOQVSTEFVEKVEYLPFVIVEPTE 240

QY 241 KFYIYNEKGLVITITARFLYKKGVEGTAFVIFGIDQGEQRISLPESLKRIPIDGSGEV 300

Db 241 KFYIYNEKGLVITITARFLYKKGVEGTAFVIFGIDQGEQRISLPESLKRIPIDGSGEV 300

QY 301 VLSRKLVDGQVQNPRAEDLVGKSLVYSATVILHSGSDMVQAESGIPITVSPYQIHFTKT 360

Db 301 VLSRKLVDGQVQNPRAEDLVGKSLVYSATVILHSGSDMVQAESGIPITVSPYQIHFTKT 360

QY 361 PKYFKGMPFDMVFTVTPNPDGSPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSQKPL 420

Db 361 PKYFKGMPFDMVFTVTPNPDGSPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSQKPL 420

QY 421 SITVTKKQELSEAEQATRTMQALPYSTVGNNSNYLHLSVLTETLRPGETLNVPFLRMD 480

Db 421 SITVTKKQELSEAEQATRTMQALPYSTVGNNSNYLHLSVLTETLRPGETLNVPFLRMD 480

QY 481 RAHEAKIRYTYLIMNKGRLKAGROVRPQGDVLVPLSITTDIPFRLVAYITLIGA 540

Db 481 RAHEAKIRYTYLIMNKGRLKAGROVRPQGDVLVPLSITTDIPFRLVAYITLIGA 540

QY 541 SGQREVADSVWVDKSDCVGSLVWKSQSEDRQVPVQGMTLKIEGHGARGVVLVAVDK 600

Db 541 SGQREVADSVWVDKSDCVGSLVWKSQSEDRQVPVQGMTLKIEGHGARGVVLVAVDK 600

QY 601 GVFLNKKKLTOSKIDWVVEKADICTPGSKDYGVSAGITFTSSSQQAQRAEL 660

Db 601 GVFLNKKKLTOSKIDWVVEKADICTPGSKDYGVSAGITFTSSSQQAQRAEL 660

QY 661 QCQPAARRRRSVQLTETKMDKVKYKPELRKCCEDGMRENPMRFSQCRRTFRFISLGEAC 720

Db 661 QCQPAARRRRSVQLTETKMDKVKYKPELRKCCEDGMRENPMRFSQCRRTFRFISLGEAC 720

QY 721 KKVFLDCCNVITELRQCHARASHGLIARSLNLDDEIIAENIVSRSEFFESLWNVEDLKE 780

Db 721 KAVFLDCCNYITELRQARASHGLARSLNDDIIEAENIVSRSEFFSWLWVNDLKE 780  
Qy 781 PPKNGISTKLMNIFLKOSITTWELAVMSDKGICVADPFVTVNQDFFIDLRLPYSV 840  
Db 781 PPKNGISTKLMNIFLKOSITTWELAVMSDKGICVADPFVTVNQDFFIDLRLPYSV 840  
Qy 841 RNEQVEIRAVLYNRQNELKVRVELLHNPAFCSLATTKERHOOTITIPKSSLSVPYV 900  
Db 841 RNEQVEIRAVLYNRQNELKVRVELLHNPAFCSLATTKERHOOTITIPKSSLSVPYV 900  
Qy 901 VPLKTGLQEVKAAVYHFTISDGVRSKLVKVPPEGIRMNKTAVRTLDPERLGRGVQKE 960  
Db 901 VPLKTGLQEVKAAVYHFTISDGVRSKLVKVPPEGIRMNKTAVRTLDPERLGRGVQKE 960  
Qy 961 DIPPADLSQVDPDTESETRILLQGTTPVAQMTEDAVDAERLKLIVTPSGCGEONMIGTP 1020  
Db 961 DIPPADLSQVDPDTESETRILLQGTTPVAQMTEDAVDAERLKLIVTPSGCGEONMIGTP 1020  
Qy 1021 TVIAVHYLDETEQWKEGLEKROGALSLIKKGYTOQLAPROPSAFAAVKAPSTWLT 1080  
Db 1021 TVIAVHYLDETEQWKEGLEKROGALSLIKKGYTOQLAPROPSAFAAVKAPSTWLT 1080  
Qy 1081 YVVKVFLAVNLIAIDSVLCGAVKWLILEKQKPDGVFOEDAPVHQMIGILRNNEKD 1140  
Db 1081 YVVKVFLAVNLIAIDSVLCGAVKWLILEKQKPDGVFOEDAPVHQMIGILRNNEKD 1140  
Qy 1141 MALTAFLVLSIQEAKDICEQVNSLPGSITKAGDPLEANVMILORSYTVIAIAGYALQNG 1200  
Db 1141 MALTAFLVLSIQEAKDICEQVNSLPGSITKAGDPLEANVMILORSYTVIAIAGYALQNG 1200  
Qy 1201 RLKGPLLNKFTITAKDNWEDPGKQLYNYEATSYALLALLQKDFPVPVVRMLNEQR 1260  
Db 1201 RLKGPLLNKFTITAKDNWEDPGKQLYNYEATSYALLALLQKDFPVPVVRMLNEQR 1260  
Qy 1261 YGGYGGYSTQITFWFQALAYQKADPHQELNLDVSLQPLSPSSKITHRIHWEASLLR 1320  
Db 1261 YGGYGGYSTQITFWFQALAYQKADPHQELNLDVSLQPLSPSSKITHRIHWEASLLR 1320  
Qy 1321 SEETKENGFTVTAEGKGQGLSVVTVYHAKADQLTCNKFDLKVTIKPATEKRPQDA 1380  
Db 1321 SEETKENGFTVTAEGKGQGLSVVTVYHAKADQLTCNKFDLKVTIKPATEKRPQDA 1380  
Qy 1381 KNTMLEICTRYRGDQATMSILDISMTGFPDPTDLKOLANGVDVRYISKVELDKAFSD 1440  
Db 1381 KNTMLEICTRYRGDQATMSILDISMTGFPDPTDLKOLANGVDVRYISKVELDKAFSD 1440  
Qy 1441 RNTLIYLDKVSHEDDCLAFKHQYFNVELIQGAVKVYAYVYNNLEESCTRFYHPEKSDG 1500  
Db 1441 RNTLIYLDKVSHEDDCLAFKHQYFNVELIQGAVKVYAYVYNNLEESCTRFYHPEKSDG 1500  
Qy 1501 KLNKLCRDELRCRAEENCFOKSDDKVTLEERLDKACEPGVDVYVYKTRLVKQLSNDPDE 1560  
Db 1501 KLNKLCRDELRCRAEENCFOKSDDKVTLEERLDKACEPGVDVYVYKTRLVKQLSNDPDE 1560  
Qy 1561 YMAIEQITKSGSEVQVQOQRTFISPKREALKLEKHVLMWGLSDFGWGKPNLSY 1620  
Db 1561 YMAIEQITKSGSEVQVQOQRTFISPKREALKLEKHVLMWGLSDFGWGKPNLSY 1620  
Qy 1621 IIGKDTWVHWPEDDECOENQKQCDLGAFTESMVVFGCPN 1663  
Db 1621 IIGKDTWVHWPEDDECOENQKQCDLGAFTESMVVFGCPN 1663

## RESULT 12

ADD93518

ID ADD93518 standard; protein; 1663 AA.

XX AC

XX ADD93518;

XX 29-JAN-2004 (first entry)

XX DT

XX Novel NOV1a, homologous to human complement C3 precursor.

DE

XX NOV1a; human; complement C3; gene therapy.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX Peptide 1..22  
XX Protein /note="Signal peptide"  
XX /note="Maure protein"  
XX W02003078572-A2.  
XX 25-SEP-2003.  
XX 06-MAR-2003; 2003WO-US006859.  
XX 15-MAR-2002; 2002US-0365034P.  
XX 19-MAR-2002; 2002US-036547P.  
XX 21-MAR-2002; 2002US-0366420P.  
XX 05-MAR-2003; 2003US-00379747.  
XX (CURA-) CURAGEN CORP.  
XX Burgess CE, Chant JS, Chaudhuri A, Edinger SR, Gangolli EA;  
XX Malyankar UM, Miller CE, Ooi CE, Ort T, Patturajan M, Rastelli L;  
XX Rieger DK, Shimkets RA, Zehnhusen BD;  
XX WPI; 2003-779122/73.  
XX N-PSDB; ADD93517.

New isolated NOVX polypeptides and polynucleotides, useful for

preventing, diagnosing or treating NOVX-associated disorders, e.g. osteoarthritis, obesity, atherosclerosis, cancer, parkinson's disease, asthma, or infections.

Claim 1; Page 100; 205pp; English.

The present sequence is the protein sequence of a novel polypeptide, designated NOV1a, that shows 99% sequence homology to the human complement C3 precursor. The invention is based on the identification of proteins and polypeptides, and the nucleic acids encoding them, that are differentially modulated in a pathological state, disease or an abnormal condition or state. These are targets for therapeutic agents and can be used in screening methodologies to identify candidate therapeutic agents which interact with the target and thereby exert a desired or favourable effect, e.g. in neurogenesis, cell differentiation, cell proliferation, haematopoiesis, wound healing and angiogenesis. Methods for diagnosis, treatment and prevention of disorders involving the novel human nucleic acids and proteins are provided. The polypeptides can also be used to raise antibodies useful e.g. in diagnosis and therapy.

Query Match 99.9%; Score 8598; DB 7; Length 1663;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1661; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGPTSGPSLLLLTHLPALGSPMYSIITPNILRLESEETWVLEAHDAGQDVPVTVVH 60  
Db 1 MGPTSGPSLLLLTHLPALGSPMYSIITPNILRLESEETWVLEAHDAGQDVPVTVVH 60  
Qy 61 DFGKLVLSSEKTVLTPATNMGNVTITIPANREFSEKGRNKFVTVQATFGTVVVKV 120  
Db 61 DFGKLVLSSEKTVLTPATNMGNVTITIPANREFSEKGRNKFVTVQATFGTVVVKV 120  
Qy 121 VLVSQGVLFITQDKTIYTPGSTVLYRIFTVNHLKLPVGTVMVNIENPEGIPVKQDSL 180  
Db 121 VLVSQGVLFITQDKTIYTPGSTVLYRIFTVNHLKLPVGTVMVNIENPEGIPVKQDSL 180  
Qy 181 SSONQLGVLPISWDIPELVNMGMQKIRAYENSQOVFSTFEFVKEVYLSFEVIVPEPTE 240  
Db 181 SSONQLGVLPISWDIPELVNMGMQKIRAYENSQOVFSTFEFVKEVYLSFEVIVPEPTE 240

QY 241 KFYIYNEKLEVTITAREFLYKKGVEGTAFVIFGIQDGEORISLPSLKRPIEDSGSEV 300  
DB 241 KFYIYNEKLEVTITAREFLYKKGVEGTAFVIFGIQDGEORISLPSLKRPIEDSGSEV 300  
QY 301 VLSKVLDDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQABRSQIPVTSFYQHFTKT 360  
DB 301 VLSKVLDDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQABRSQIPVTSFYQHFTKT 360  
QY 361 PKYKPCWPDVAVFVTPNDGSPAYRVPVAVOGEDEVQSLTQGDGAVKLSTNTHPSOKPL 420  
DB 361 PKYKPCWPDVAVFVTPNDGSPAYRVPVAVOGEDEVQSLTQGDGAVKLSTNTHPSOKPL 420  
QY 421 SITVTKKQELSAEQATRMQALPYSTVGNSSNNYHLVSLRTELPGGETLNVLNFLRMD 480  
DB 421 SITVTKKQELSAEQATRMQALPYSTVGNSSNNYHLVSLRTELPGGETLNVLNFLRMD 480  
QY 481 RAHEAKRYTYILMKNKGRLLKAGROVREPQDGLVPLSITTDPIPSFRILVAYYTLIGA 540  
DB 481 RAHEAKRYTYILMKNKGRLLKAGROVREPQDGLVPLSITTDPIPSFRILVAYYTLIGA 540  
QY 541 SGQREVAVDSVWVDKSCVGSILVWKSQSEDRQVPVQGMFLKIEGDHGRVAVLVADK 600  
DB 541 SGQREVAVDSVWVDKSCVGSILVWKSQSEDRQVPVQGMFLKIEGDHGRVAVLVADK 600  
QY 601 GVFLVKNKLTOSKIVDVEKADIGCTPGSGKDVAGVFSQAGLTFTSSSGOQTAQRAEL 660  
DB 601 GVFLVKNKLTOSKIVDVEKADIGCTPGSGKDVAGVFSQAGLTFTSSSGOQTAQRAEL 660  
QY 661 QCPQPAARRRSVQLTEKMDKVKYKPKELRKCCEGDMENPMRPSQORRTFSLGEAC 720  
DB 661 QCPQPAARRRSVQLTEKMDKVKYKPKELRKCCEGDMENPMRPSQORRTFSLGEAC 720  
QY 721 KKVFLDCNVTILRQHARASHLGLARSNLDEDIIAENIVSRSEFPESMLWNVDELKE 780  
DB 721 KKVFLDCNVTILRQHARASHLGLARSNLDEDIIAENIVSRSEFPESMLWNVDELKE 780  
QY 781 PPKNGISTKLMNIFLKDSTITWEILAVSDSKGICVADPPBEVTWQDFFIDLPLPSVY 840  
DB 781 PPKNGISTKLMNIFLKDSTITWEILAVSDSKGICVADPPBEVTWQDFFIDLPLPSVY 840  
QY 841 RNEQVEIRAVLYNRYQKQELKRVVELLHNPAPCSLATTKRHQQTITIPPKSSISVPIVI 900  
DB 841 RNEQVEIRAVLYNRYQKQELKRVVELLHNPAPCSLATTKRHQQTITIPPKSSISVPIVI 900  
QY 901 VPLKTGQVEVKAAYVHHFISDGVKSLKVYVEGIRMNKTVAVRTLDPERLREGVQKE 960  
DB 901 VPLKTGQVEVKAAYVHHFISDGVKSLKVYVEGIRMNKTVAVRTLDPERLREGVQKE 960  
QY 961 DIPPADLSQVDPTESETRILQGTPEVAQMTEDAVDAERLKLIVTPSGCGEQNMIGTP 1020  
DB 961 DIPPADLSQVDPTESETRILQGTPEVAQMTEDAVDAERLKLIVTPSGCGEQNMIGTP 1020  
QY 1021 TVIAVHVLDETEQWKEGKRGALIKKGTQQLAPROPSSAFAAFVRAEFTWLTA 1080  
DB 1021 TVIAVHVLDETEQWKEGKRGALIKKGTQQLAPROPSSAFAAFVRAEFTWLTA 1080  
QY 1081 YVVKVFLAVNLIAIDSQLCGAVKWLILEKQKPDGVFOEDAPVTHQEMIGLNNNEKD 1140  
DB 1081 YVVKVFLAVNLIAIDSQLCGAVKWLILEKQKPDGVFOEDAPVTHQEMIGLNNNEKD 1140  
QY 1141 MALTAFLVLSIQAKDICEBOVNSLPGSITKAGDFLEANYMNLORSYTVAGVALAQMG 1200  
DB 1141 MALTAFLVLSIQAKDICEBOVNSLPGSITKAGDFLEANYMNLORSYTVAGVALAQMG 1200  
QY 1201 RLKGFPLNKLFTAKQKRWEDSGKQLYNVEATSTALLALQLKDFDVPVPPVRLNEQR 1260  
DB 1201 RLKGFPLNKLFTAKQKRWEDSGKQLYNVEATSTALLALQLKDFDVPVPPVRLNEQR 1260  
QY 1261 YVGGYGSTQATFWFOALAQYQKQADPQHELNLDVLSQLPSRSSKITHRIHWSASILLR 1320  
DB 1261 YVGGYGSTQATFWFOALAQYQKQADPQHELNLDVLSQLPSRSSKITHRIHWSASILLR 1320

QY 1321 SEETKENEGETVTAEGKGQGLTSVWTMYHAKAKDQLTCKNFKDKVTIKPAPETEKRPQDA 1380  
DB 1321 SEETKENEGETVTAEGKGQGLTSVWTMYHAKAKDQLTCKNFKDKVTIKPAPETEKRPQDA 1380  
QY 1381 KNTWILICTRYRGDQDATMSILDISMTGFPAPDTDDQLKOLANGVDRVYISKYELDKAFSD 1440  
DB 1381 KNTWILICTRYRGDQDATMSILDISMTGFPAPDTDDQLKOLANGVDRVYISKYELDKAFSD 1440  
QY 1441 RNTLIIYLDKVSSEDDCLAFKVOYFNVLEIQPGAVKYVAYNLEESCTRYHPEKEDG 1500  
DB 1441 RNTLIIYLDKVSSEDDCLAFKVOYFNVLEIQPGAVKYVAYNLEESCTRYHPEKEDG 1500  
QY 1501 KLNKCRDELRCRAEENCFIQKSDDKVTLEELDKACBEGVDYVYKTLVKVQLSNDEDE 1560  
DB 1501 KLNKCRDELRCRAEENCFIQKSDDKVTLEELDKACBEGVDYVYKTLVKVQLSNDEDE 1560  
QY 1561 YIMAEOTIKSGSDEVQVGQORTFISPIKCREALKLEKKHYLMWGLSSDFWGEKPNLSY 1620  
DB 1561 YIMAEOTIKSGSDEVQVGQORTFISPIKCREALKLEKKHYLMWGLSSDFWGEKPNLSY 1620  
QY 1621 IIGKDTVWEHWPBEDECOQDEENKQCODLGAFTESMVVFGCPN 1663  
DB 1621 IIGKDTVWEHWPBEDECOQDEENKQCODLGAFTESMVVFGCPN 1663

## RESULT 13

ADK12322  
ID ADK12322 standard; protein; 1663 AA.

XX AC ADK12322;

XX DT 20-MAY-2004 (first entry)

XX XX Human complement component C3.

XX KW Antisense therapy; human; complement component C3; autoimmune disorder;  
multiple sclerosis; infection; atherosclerosis; neuroprotective;  
XX KW antiatherosclerotic; antimicrobial; antiinflammatory; cytostatic.

XX OS Homo sapiens.

XX PN US2004043956-A1.

XX PD 04-MAR-2004.

XX PF 18-AUG-2003; 2003US-00642802.

XX PR 23-OCT-2001; 2001US-00001076.

XX XX (GRAH/) GRAHAM M J.

XX PA (WATT/) WATT A T.

XX XX Graham MJ, Watt AT;

XX XX WPI: 2004-225730/21.

XX DR N-PSDS; ADK12146.

XX XX New antisense compound targeted to a nucleic acid molecule encoding  
complement component C3, useful for treating multiple sclerosis, an  
infection or atherosclerosis.

XX PS Disclosure; Page 31-36; 74pp; English.

XX CC The present invention relates to antisense compounds targeted to a  
nucleic acids encoding human and mouse complement component C3. The  
antisense compound comprises an antisense oligonucleotide that  
specifically hybridizes with the nucleic acid and inhibits the expression  
of complement component C3 in cells. The antisense oligonucleotide is a  
chimeric oligonucleotide. The antisense oligonucleotide comprises at  
least one modified internucleoside linkage, preferably a phosphorothioate  
linkage. It also comprises at least one modified sugar moiety, preferably  
a 2'-O-methoxyethyl (2'-MOE) sugar moiety. The antisense oligonucleotide  
further comprises at least one modified nucleobase, preferably a 5-

CC methionylcysteine. The antisense oligonucleotides are useful for the  
 CC treatment of diseases such as autoimmune disorders e.g. multiple  
 CC sclerosis, infections, and atherosclerosis. The present sequence  
 CC represents human complement component C3.  
 XX  
 SQ Sequence 1663 AA;

Query Match 99.9%; Score 8598; DB 8; Length 1663;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1663; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGPTSGSLLLLLLTHPLALGSPMYSIITPNILRLESEETWVLEAHDAQSDVPVTVH 60  
 Db 1 MGPTSGSLLLLLLTHPLALGSPMYSIITPNILRLESEETWVLEAHDAQSDVPVTVH 60

Qy 61 DFGKGLVLSEKTLVTPATNHMGNTFTIPANBEKSEKGNKFTVQACFTGTOVBEK 120  
 Db 61 DFGKGLVLSEKTLVTPATNHMGNTFTIPANBEKSEKGNKFTVQACFTGTOVBEK 120

Qy 121 VLVSQGYLFIQTDKTIYTPGSTVLYRFTVNHKLLPVGRVVMNINPESGIPVKQDSL 180  
 Db 121 VLVSQGYLFIQTDKTIYTPGSTVLYRFTVNHKLLPVGRVVMNINPESGIPVKQDSL 180

Qy 181 SSNQLGVPLSWDIPELVNMGWKIRAYENSPOQVSTEFVEKYYLPSFEVIVPE 240  
 Db 181 SSNQLGVPLSWDIPELVNMGWKIRAYENSPOQVSTEFVEKYYLPSFEVIVPE 240

Qy 241 KFYIYNEKLEVTITARFLYKKGVEGTAFVFGIQDGEQRSLPESLKRPIEDGSGEV 300  
 Db 241 KFYIYNEKLEVTITARFLYKKGVEGTAFVFGIQDGEQRSLPESLKRPIEDGSGEV 300

Qy 301 VLGRKVLDDGVQNPRAEDLVGKSLYSATVILHSGSDMVQABRSIGPIVTPYQHFTK 360  
 Db 301 VLGRKVLDDGVQNPRAEDLVGKSLYSATVILHSGSDMVQABRSIGPIVTPYQHFTK 360

Qy 361 PKYFKQMPDLVFTVTPDGSPPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSKPL 420  
 Db 361 PKYFKQMPDLVFTVTPDGSPPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSKPL 420

Qy 421 SITVTKKQLESEABQATRMQALPSTVGNSSNNYLHLSVLTARPGETLNVNPLLRMD 480  
 Db 421 SITVTKKQLESEABQATRMQALPSTVGNSSNNYLHLSVLTARPGETLNVNPLLRMD 480

Qy 481 RAHEAKIRYTYILMKGRLKAGROVRPQDQVPLVLSITTDIPFRLVAYTYLIGA 540  
 Db 481 RAHEAKIRYTYILMKGRLKAGROVRPQDQVPLVLSITTDIPFRLVAYTYLIGA 540

Qy 541 SGQREVADSVWVDVKDSCVGLVWKSQSEDRQPVPGQMTLKEGHDGARVILVADK 600  
 Db 541 SGQREVADSVWVDVKDSCVGLVWKSQSEDRQPVPGQMTLKEGHDGARVILVADK 600

Qy 601 GVFLNKKKLTQSKIMDVVEKADIGCTPGSGKDYAGVPSDAGLTFTSSSQQTARAE 660  
 Db 601 GVFLNKKKLTQSKIMDVVEKADIGCTPGSGKDYAGVPSDAGLTFTSSSQQTARAE 660

Qy 661 QCPOPARRRRSVQLTEKMDKVGKYPKELRKCCEDGMENPMRPSQORRTFISLGAC 720  
 Db 661 QCPOPARRRRSVQLTEKMDKVGKYPKELRKCCEDGMENPMRPSQORRTFISLGAC 720

Qy 721 KKVFLDCNCTITELRQARASHGLARSNLDEIIAENIVSRSEFFESMWNVEDLKE 780  
 Db 721 KKVFLDCNCTITELRQARASHGLARSNLDEIIAENIVSRSEFFESMWNVEDLKE 780

Qy 781 PPKNIGISTKLMNIFLKDSTITWEILAVSMDSKGIQVADPFVTVNQDFFIDRLRPSVW 840  
 Db 781 PPKNIGISTKLMNIFLKDSTITWEILAVSMDSKGIQVADPFVTVNQDFFIDRLRPSVW 840

Qy 841 RNEQVEIRAVLYNRQNEQLKRVVLELHNPAPFCSLATTGRHCQOTITPPKSSLSVPVYI 900  
 Db 841 RNEQVEIRAVLYNRQNEQLKRVVLELHNPAPFCSLATTGRHCQOTITPPKSSLSVPVYI 900

Qy 901 VPLTKGLOEVEVKAAYVHHFISDGVKSLKVVPEGIRMNKTVAVRTLDPERLREGVQKE 960  
 Db 901 VPLTKGLOEVEVKAAYVHHFISDGVKSLKVVPEGIRMNKTVAVRTLDPERLREGVQKE 960

## RESULT 14

ADN04780

ID ADN04780 standard; protein; 1663 AA.

XX AC ADN04780;

XX AC ADN04780;

XX DT 01-JUL-2004 (first entry)

XX DE Antisporiatric protein sequence #570.

XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.

XX OS Homo sapiens.

XX XX WO2004028479-A2.

XX PD 08-APR-2004.

XX XX 25-SEP-2003; 2003WO-US030907.

XX XX 25-SEP-2002; 2002US-0414006P.

XX XX

PA (GETH) GENENTECH INC.  
 XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WT;  
 PI Wu TD;  
 XX WPI; 2004-305105/28.  
 DR N-PSDB; ADNR04779.  
 XX New PRO nucleic acid or polypeptide, useful for preparing a  
 PT pharmaceutical composition for diagnosing or treating psoriasis in a  
 PT mammal.  
 XX Claim 9; SEQ ID NO 1174; 3069pp; English.  
 XX The invention relates to novel polynucleotide and polypeptides for  
 CC treating psoriasis or a sequence having at least 80% identity to the  
 CC above sequences. The nucleic acid is useful for preparing a composition  
 CC for diagnosing or treating psoriasis in a mammal. This sequence  
 CC corresponds to one of the polypeptides of the invention.  
 XX Sequence 1663 AA;  
 SQ

Query Match 99.9%; Score 8598; DB 8; Length 1663;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1661; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGPTSGPSLLLLTHLPLALGSPMYSIITPNILRLESEETMVLHAHQDQGVPTVTVH 60  
 Db 1 MGPTSGPSLLLLTHLPLALGSPMYSIITPNILRLESEETMVLHAHQDQGVPTVTVH 60  
 Qy 61 DPEGKLVLSSEKTVLTPATNMGVNTFTIPANREFKSEKGRNFVTVQATFGVQVEKV 120  
 Db 61 DPEGKLVLSSEKTVLTPATNMGVNTFTIPANREFKSEKGRNFVTVQATFGVQVEKV 120  
 Qy 121 VLVSQSGYLFQTDKTIYTPGSTVLYRIFTVNHKLLFVGRVTVMVNIENPEGIPVKQDSL 180  
 Db 121 VLVSQSGYLFQTDKTIYTPGSTVLYRIFTVNHKLLFVGRVTVMVNIENPEGIPVKQDSL 180  
 Qy 181 SSONQGLVPLNDIPELNMGQWIRAYYENSPOOVSTEFVEKVEYVLPSEFVIEPTE 240  
 Db 181 SSONQGLVPLNDIPELNMGQWIRAYYENSPOOVSTEFVEKVEYVLPSEFVIEPTE 240  
 Qy 241 KFYIYNKGLVITITARELYGKVEGTAFVIFGIDQGEQISLPESLKRIPEDSGSEV 300  
 Db 241 KFYIYNKGLVITITARELYGKVEGTAFVIFGIDQGEQISLPESLKRIPEDSGSEV 300  
 Qy 301 VLSRKVLGVQNPRAEDLVGSLVSVATVILHSGSDMVQAESGIPITVSPYQIHFTKT 360  
 Db 301 VLSRKVLGVQNPRAEDLVGSLVSVATVILHSGSDMVQAESGIPITVSPYQIHFTKT 360  
 Qy 361 PKYFKGMPFDMVFTNPDGSPAYRVPVAVQGEDTVQSLTQGGGVAKLSINTHPSQKPL 420  
 Db 361 PKYFKGMPFDMVFTNPDGSPAYRVPVAVQGEDTVQSLTQGGGVAKLSINTHPSQKPL 420  
 Qy 421 SIIVRTKQELSEAEQATMQLPVSTVGRSNNVHLVLTRELPGETLNNFLRMD 480  
 Db 421 SIIVRTKQELSEAEQATMQLPVSTVGRSNNVHLVLTRELPGETLNNFLRMD 480  
 Qy 481 RAHEAKIRYTYLIMNKGRLKAGQVREPQGLVPLSITTDPIPSFLVAYYTLIGA 540  
 Db 481 RAHEAKIRYTYLIMNKGRLKAGQVREPQGLVPLSITTDPIPSFLVAYYTLIGA 540  
 Qy 541 SGQREVVDVSVVDVSDCVGSLVVKSGQSDRQVPVQCOQMTLKEGDHGRVVLVAVDK 600  
 Db 541 SGQREVVDVSVVDVSDCVGSLVVKSGQSDRQVPVQCOQMTLKEGDHGRVVLVAVDK 600  
 Qy 601 GVFLVNNKNTQSKLWDVVEKADICTTPGSGKVAGVFSADGLTFTSSSGQQTAQRAEL 660  
 Db 601 GVFLVNNKNTQSKLWDVVEKADICTTPGSGKVAGVFSADGLTFTSSSGQQTAQRAEL 660  
 Qy 661 QCQPPAARRRRSVQLTKEKMDVGVKYPKELRKCCEGDMRENPMRFSQORTRFTSLGEAC 720  
 Db 661 QCQPPAARRRRSVQLTKEKMDVGVKYPKELRKCCEGDMRENPMRFSQORTRFTSLGEAC 720

Qy 721 KKVFLDCNCVITELRRQHARASHLGLARSNLDEBIIAENIVSRSEFFESMWNVEDLKE 780  
 Db 721 KKVFLDCNCVITELRRQHARASHLGLARSNLDEBIIAENIVSRSEFFESMWNVEDLKE 780  
 Qy 781 PPKNGISTKLMNIFLKDSITTWELAVSDSKGICVADPFEVTVMQDFFIDRLRPSYV 840  
 Db 781 PPKNGISTKLMNIFLKDSITTWELAVSDSKGICVADPFEVTVMQDFFIDRLRPSYV 840  
 Qy 841 RNEQVEIRAVLYNRYRQNELKVRVELLHNPAPFCSLATTKRHOQTITIPPKSSLSVPYV 900  
 Db 841 RNEQVEIRAVLYNRYRQNELKVRVELLHNPAPFCSLATTKRHOQTITIPPKSSLSVPYV 900  
 Qy 901 VPLTGTQEVVEKAAVYVHFIISGVKSLKVPEGRMNKTVAVRTLPERLGRGVQKE 960  
 Db 901 VPLTGTQEVVEKAAVYVHFIISGVKSLKVPEGRMNKTVAVRTLPERLGRGVQKE 960  
 Qy 961 DIPPADLSQVPTDTESTRILLQGTTPVAQMTEDAVDAERLKLIVTPSGCGEQNMIGWTP 1020  
 Db 961 DIPPADLSQVPTDTESTRILLQGTTPVAQMTEDAVDAERLKLIVTPSGCGEQNMIGWTP 1020  
 Qy 1021 TVIAVHYLDTEQWEKFGLEKROGALIELIKGYTQOLAFRQPSAFAAFVKEAPSTWLTA 1080  
 Db 1021 TVIAVHYLDTEQWEKFGLEKROGALIELIKGYTQOLAFRQPSAFAAFVKEAPSTWLTA 1080  
 Qy 1081 YVVKVFLAVNLIAIDSVQLCGAVKWLILEKQKPDGVFOEDAPVTHQEMIGGLRNNKED 1140  
 Db 1081 YVVKVFLAVNLIAIDSVQLCGAVKWLILEKQKPDGVFOEDAPVTHQEMIGGLRNNKED 1140  
 Qy 1141 MALTAFLVLSIQAKDICEBOVNSLPGSITKAGDFLEANYMNLQRSYTVAIAGYALAQWG 1200  
 Db 1141 MALTAFLVLSIQAKDICEBOVNSLPGSITKAGDFLEANYMNLQRSYTVAIAGYALAQWG 1200  
 Qy 1201 RLKGPLNKELITTAIDKKNRWEDEPKQLYNVEATSYALLALLQKDFDVPVVRWLNQSR 1260  
 Db 1201 RLKGPLNKELITTAIDKKNRWEDEPKQLYNVEATSYALLALLQKDFDVPVVRWLNQSR 1260  
 Qy 1261 YVGGYGSTQATFWFQALAQYQKADPADHQLNLDVSLQPSRSSKITTHRIHWESASLLR 1320  
 Db 1261 YVGGYGSTQATFWFQALAQYQKADPADHQLNLDVSLQPSRSSKITTHRIHWESASLLR 1320  
 Qy 1321 SEETKENEFTVTAEGKGGQTLVTVWYHAKAKDQLTCKNFKLVITIKAPATEKRPQDA 1380  
 Db 1321 SEETKENEFTVTAEGKGGQTLVTVWYHAKAKDQLTCKNFKLVITIKAPATEKRPQDA 1380  
 Qy 1381 KNTMILBICRYRGDQDATMSILLDISMTGFAPDTDDLQKLANGVDVRIKVELDKAFSD 1440  
 Db 1381 KNTMILBICRYRGDQDATMSILLDISMTGFAPDTDDLQKLANGVDVRIKVELDKAFSD 1440  
 Qy 1441 RNTLIIYLDKVSSEDDCLAFKHQVFNVELTOPGAKVYAYNLEESCTRYHPEKEDG 1500  
 Db 1441 RNTLIIYLDKVSSEDDCLAFKHQVFNVELTOPGAKVYAYNLEESCTRYHPEKEDG 1500  
 Qy 1501 KLNKLCRDELRCRAEENCFIQKSDDKVTLLEERLDKACEPGVDVYKTRLVKQLSNDEDE 1560  
 Db 1501 KLNKLCRDELRCRAEENCFIQKSDDKVTLLEERLDKACEPGVDVYKTRLVKQLSNDEDE 1560  
 Qy 1561 YMAIEQTIKSGSEDEVQVGGQRTFIPKICREALKLEBKHYLMWGLSSDFWGEKPNLSY 1620  
 Db 1561 YMAIEQTIKSGSEDEVQVGGQRTFIPKICREALKLEBKHYLMWGLSSDFWGEKPNLSY 1620  
 Qy 1621 IIGKDTVWEHWPEDQECQDEENKQCDLGAFTESMVVFGCFN 1663  
 Db 1621 IIGKDTVWEHWPEDQECQDEENKQCDLGAFTESMVVFGCFN 1663

RESULT 15  
 AAW40988  
 ID AAW40988 standard; protein; 1663 AA.  
 XX  
 AC AAW40988;  
 XX  
 DT 09-APR-1998 (first entry)



QY 1201 RLKGPLINKELTTAKDKWNEDEPGKOLYNVENTSYALLALLQLKDFDPVPPVWVWNEQR 1260  
DB 1201 RLKGPLINKELTTAKDKWNEDEPGKOLYNVENTSYALLALLQLKDFDPVPPVWVWNEQR 1260  
QY 1261 YGGGYGSTOATFMVFCALAQKQAPDHQELNDVSLQLPFSRSKITHRIHWESASLLR 1320  
DB 1261 YGGGYGSTOATFMVFCALAQKQAPDHQELNDVSLQLPFSRSKITHRIHWESASLLX 1320  
QY 1321 SEETKENEGFTVTAEGKGOGTISVVTWVHAKAKDQJTCNKFDLKVTIKPAFETEKRPQDA 1380  
DB 1321 SEETKENEGFTVTAEGKGOGTISVVTWVHAKAKDQJTCNKFDLKVTIKPAFETEKRPQDA 1380  
QY 1381 KNTMILEICTRYRGDQDATMSILDISMMTGPAPDQDDLLKQLANGVDRYISKYELDKAFSD 1440  
DB 1381 KNTMILEICTRYRGDQDATMSILDISMMTGPAPDQDDLLKQLANGVDRYISKYELDKAFSD 1440  
QY 1441 RNTLIYLDKVSHSEDDCLAFKVHOFVNVVELIQGAVKVYAYYNLEBSCTRFYHPEKEDG 1500  
DB 1441 RNTLIYLDKVSHSEDDCLAFKVHOFVNVVELIQGAVKVYAYYNLEBSCTRFYHPEKEDG 1500  
QY 1501 KLKLCRDELCRCABENCTIQKSDDKVTLLEERLDKACBPGVDYVYKTLVQVLSNDFDE 1560  
DB 1501 KLKLCRDELCRCABENCTIQKSDDKVTLLEERLDKACBPGVDYVYKTLVQVLSNDFDE 1560  
QY 1561 YIWAIEOTIKSGSDEVQVQOQRTFISPIKCREALKLEEKHYLMWGLSSDFWGEKPNLSY 1620  
DB 1561 YIWAIEOTIKSGSDEVQVQOQRTFISPIKCREALKLEEKHYLMWGLSSDFWGEKPNLSY 1620  
QY 1621 IIGKDTWVHWPDEECQDEENQKQCDLGAFTESNVVFGCPN 1663  
DB 1621 IIGKDTWVHWPDEECQDEENQKQCDLGAFTESNVVFGCPN 1663

Search completed: December 22, 2004, 00:33:06  
Job time : 205 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2004, 00:33:14 ; Search time 191 Seconds  
(without alignments)  
3115.599 Million cell updates/sec

Title: US-09-875-519A-22  
Perfect score: 8609  
Sequence: 1 MGPTSGPSLLLLLLTHPLA.....KQCQDLGARTSNVVGCPN 1663

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1589859 seqs, 357834939 residues

Total number of hits satisfying chosen parameters: 1589859

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8609	100.0	1663	9	US-09-875-519A-22
2	8598	99.9	1663	10	US-09-842-758-41
3	8598	99.9	1663	15	US-10-379-747-2
4	8598	99.9	1663	15	US-10-174-333-41
5	6752	78.4	1663	10	US-09-842-758-43
6	6752	78.4	1663	15	US-10-174-333-43
7	6724	78.1	1661	10	US-09-842-758-42
8	6724	78.1	1661	15	US-10-174-333-42
9	4224	48.1	1642	9	US-09-925-442-2
10	4194	48.7	1648	9	US-09-925-442-35
11	3536	41.1	1333	9	US-09-925-442-34
12	3260	37.9	705	15	US-10-379-747-4
13	2857.5	33.2	1612	10	US-09-842-758-6

14	2857.5	33.2	1612	15	US-10-174-333-6	Sequence 6, Appl
15	1855.5	21.6	751	10	US-09-981-151A-80	Sequence 80, Appl
16	1855.5	21.6	751	15	US-10-037-417-115	Sequence 115, Appl
17	1818.5	21.1	740	11	US-09-972-211-70	Sequence 70, Appl
18	1818.5	21.1	740	15	US-10-086-625-70	Sequence 70, Appl
19	1817	21.1	1680	17	US-10-724-562-2	Sequence 2, Appl
20	1795.5	20.4	1602	9	US-09-778-927A-59	Sequence 59, Appl
21	1582	18.4	310	11	US-09-834-309-7	Sequence 7, Appl
22	1576	18.3	310	11	US-09-834-309-8	Sequence 8, Appl
23	1561	18.1	310	11	US-09-834-309-9	Sequence 9, Appl
24	1458	16.9	296	15	US-10-398-916-29	Sequence 29, Appl
25	1458	16.9	296	15	US-10-398-916-30	Sequence 30, Appl
26	1430	16.6	300	15	US-10-398-916-13	Sequence 13, Appl
27	1347.5	15.7	1251	9	US-09-778-927A-58	Sequence 58, Appl
28	1330	15.4	300	15	US-10-398-916-11	Sequence 11, Appl
29	1305	15.2	409	16	US-10-486-655-6	Sequence 6, Appl
30	1299	15.1	620	10	US-09-981-151A-81	Sequence 81, Appl
31	1299	15.1	620	15	US-10-037-417-116	Sequence 116, Appl
32	1293	15.0	312	15	US-10-398-916-17	Sequence 17, Appl
33	1266	14.7	610	11	US-09-972-211-71	Sequence 71, Appl
34	1266	14.7	610	15	US-10-096-625-71	Sequence 71, Appl
35	1265	14.7	300	15	US-10-398-916-15	Sequence 15, Appl
36	1257	14.6	300	15	US-10-398-916-9	Sequence 9, Appl
37	1017	11.8	1474	9	US-09-875-519A-22	Sequence 5, Appl
38	1017	11.8	1474	11	US-09-972-211-56	Sequence 56, Appl
39	1017	11.8	1474	11	US-09-750-972-5	Sequence 5, Appl
40	1017	11.8	1474	14	US-10-292-081A-10	Sequence 10, Appl
41	1017	11.8	1474	14	US-10-292-081A-12	Sequence 12, Appl
42	1017	11.8	1474	14	US-10-292-081A-13	Sequence 13, Appl
43	1017	11.8	1474	14	US-10-170-385-405	Sequence 405, Appl
44	1017	11.8	1474	14	US-10-331-496A-38	Sequence 38, Appl
45	1017	11.8	1474	15	US-10-608-397-10	Sequence 10, Appl

## ALIGNMENTS

RESULT 1  
US-09-875-519A-22  
; Sequence 22: Application US/09875519A  
; Patent No. US20020068059A1  
; GENERAL INFORMATION:  
; APPLICANT: Faries, Timothy C.  
; APPLICANT: Harrison, Richard A.  
; TITLE OF INVENTION: Down-Regulation Resistant C3 Convertase  
; FILE REFERENCE: 4-30443/A/IMU/PCT  
; CURRENT APPLICATION NUMBER: US/09/875,519A  
; CURRENT FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: PCT/GB97/00603  
; PRIOR FILING DATE: 1997-03-04  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1663  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-875-519A-22

Query Match 100.0%; Score 8609; DB 9; Length 1663;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGPTSGPSLLLLLLTHPLALGSPMYSIITPNILRESEETWVLEAHDAQGDVPTVTVH	60
Db	1	MGPTSGPSLLLLLLTHPLALGSPMYSIITPNILRESEETWVLEAHDAQGDVPTVTVH	60
Qy	61	DFPGKLVLSSEKTVLPATNMGNVTIIPANREFKSEKGRNFVTVQATFGGVVEKV	120
Db	61	DFPGKLVLSSEKTVLPATNMGNVTIIPANREFKSEKGRNFVTVQATFGGVVEKV	120
Qy	121	VLVSLQSGVLFITQDXTIYTPGTVLYRIFTVNHKLLPVGRTVMVNIENPESGIPVKQDSL	180
Db	121	VLVSLQSGVLFITQDXTIYTPGTVLYRIFTVNHKLLPVGRTVMVNIENPESGIPVKQDSL	180



; PRIOR APPLICATION NUMBER: 60/201,186  
 ; PRIOR FILING DATE: 2000-05-02  
 ; PRIOR APPLICATION NUMBER: 60/201,474  
 ; PRIOR FILING DATE: 2000-05-03  
 ; PRIOR APPLICATION NUMBER: 60/201,508  
 ; PRIOR FILING DATE: 2000-05-03  
 ; PRIOR APPLICATION NUMBER: 60/220,591  
 ; PRIOR FILING DATE: 2000-07-25  
 ; PRIOR APPLICATION NUMBER: 60/232,678  
 ; PRIOR FILING DATE: 2000-09-15  
 ; PRIOR APPLICATION NUMBER: 60/263,217  
 ; PRIOR FILING DATE: 2001-01-22  
 ; PRIOR APPLICATION NUMBER: 60/265,160  
 ; PRIOR FILING DATE: 2001-01-30  
 ; NUMBER OF SEQ ID NOS: 113  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 41  
 ; LENGTH: 1663  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-842-758-41

Query Match 99.9%; Score 8599; DB 10; Length 1663;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1661; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	MGPTSGPSLLLLTHPLALGSPMYSIITPNILRLSEETVLEAHDAQGGVPVTVVH	60
DB	1	MGPTSGPSLLLLTHPLALGSPMYSIITPNILRLSEETVLEAHDAQGGVPVTVVH	60
QY	61	DFPGKLVLSSEKTVLTPATNMGNTFTIPANREFKSEKRNKFTVQATFGTQVREK	120
DB	61	DFPGKLVLSSEKTVLTPATNMGNTFTIPANREFKSEKRNKFTVQATFGTQVREK	120
QY	121	VLVLSQGYLFICDXTIYTPGSTVLYRLFTVNHKLLPVGRVTWNINPEGIPVKQDSL	180
DB	121	VLVLSQGYLFICDXTIYTPGSTVLYRLFTVNHKLLPVGRVTWNINPEGIPVKQDSL	180
QY	181	SSQNQLGVPLSDWIDPELVNMGQWKIRAYENSPOQVSTEFVEKVEYVLPSEVIVEPTE	240
DB	181	SSQNQLGVPLSDWIDPELVNMGQWKIRAYENSPOQVSTEFVEKVEYVLPSEVIVEPTE	240
QY	241	KFYIYNEKLEVTITARFLYGKVGRTAFVIGIODGEOR-SLPSLKRIFEDSGSEV	300
DB	241	KFYIYNEKLEVTITARFLYGKVGRTAFVIGIODGEOR-SLPSLKRIFEDSGSEV	300
QY	301	VLSEKVLLDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQAESGIPVTSPIQHFTKT	360
DB	301	VLSEKVLLDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQAESGIPVTSPIQHFTKT	360
QY	361	PKYFKGMPDLMVFVTPNPGSPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSQPL	420
DB	361	PKYFKGMPDLMVFVTPNPGSPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSQPL	420
QY	421	SITVTRKKQELSAEQATRMQALPYSTVGNSSNNYLHLSVLRLTELRPGSETLNVNLLRMD	480
DB	421	SITVTRKKQELSAEQATRMQALPYSTVGNSSNNYLHLSVLRLTELRPGSETLNVNLLRMD	480
QY	481	RAHEAKIRYTYLIMNKGRLKAGROVRPGQDLVPLSITTDTPSPRLVAYYTLIGA	540
DB	481	RAHEAKIRYTYLIMNKGRLKAGROVRPGQDLVPLSITTDTPSPRLVAYYTLIGA	540
QY	541	SGQREVADSVVDVKDSCVGSILVXSGQSEDRQVPVGGQMTLKIEGDHGARVLVAVDK	600
DB	541	SGQREVADSVVDVKDSCVGSILVXSGQSEDRQVPVGGQMTLKIEGDHGARVLVAVDK	600
QY	601	GVFLNKNKLTOSKINDVVEKADIGCTPGSGKDYAGVFSADAGLTFTSSGQOTQRAEL	660
DB	601	GVFLNKNKLTOSKINDVVEKADIGCTPGSGKDYAGVFSADAGLTFTSSGQOTQRAEL	660
QY	661	QCPOPAARRRSVQLTEKMDKVGKYPKELRKCCEDGMRENPMRFSQCRRTTFISLGEAC	720
DB	661	QCPOPAARRRSVQLTEKMDKVGKYPKELRKCCEDGMRENPMRFSQCRRTTFISLGEAC	720

QY	721	KKVFLDCQNYITELRQCHARASHGLAPSILDEDIIABENIVSRSEFPESKLNWVEDLKE	780
DB	721	KKVFLDCQNYITELRQCHARASHGLAPSILDEDIIABENIVSRSEFPESKLNWVEDLKE	780
QY	781	PPKNGISTKLMIIFLKDSITTWELLAVSMSDKKIGICVADPEFVTVMQFFIDLRLPYSVW	840
DB	781	PPKNGISTKLMIIFLKDSITTWELLAVSMSDKKIGICVADPEFVTVMQFFIDLRLPYSVW	840
QY	841	RNEQVEIRAVLYNRYRONQELKVRVELLHNPAPCSLATTKRHHQOITITPPKSSLSVPVVI	900
DB	841	RNEQVEIRAVLYNRYRONQELKVRVELLHNPAPCSLATTKRHHQOITITPPKSSLSVPVVI	900
QY	901	VPLKTGLQEVVEKAAVYHHFISDGVKRSKLVVPEGIRMNKTVAVTLPDLREGVQKE	960
DB	901	VPLKTGLQEVVEKAAVYHHFISDGVKRSKLVVPEGIRMNKTVAVTLPDLREGVQKE	960
QY	961	DIPPADLSQVDPDTESETRILLQGTVPVQMTEDAVDAERLKHILVTPSGCGEONMIGWTP	1020
DB	961	DIPPADLSQVDPDTESETRILLQGTVPVQMTEDAVDAERLKHILVTPSGCGEONMIGWTP	1020
QY	1021	TVIAVHYLDETEQWKEFKLEKQGALELIKKGYSITQQLAFROPSSAFAAFVRAAPSTWUTA	1080
DB	1021	TVIAVHYLDETEQWKEFKLEKQGALELIKKGYSITQQLAFROPSSAFAAFVRAAPSTWUTA	1080
QY	1081	YVVKVFSLANVLIADSOVLGAVKWLILLEKQKPGVFCEDAPVTHQEMIIGGLRNNNEKD	1140
DB	1081	YVVKVFSLANVLIADSOVLGAVKWLILLEKQKPGVFCEDAPVTHQEMIIGGLRNNNEKD	1140
QY	1141	MALTAFLVLSQBAKIDICEQVNSLPGSITKAGDFLEANYMNLQSSYTVAIAGYALAQMG	1200
DB	1141	MALTAFLVLSQBAKIDICEQVNSLPGSITKAGDFLEANYMNLQSSYTVAIAGYALAQMG	1200
QY	1201	RLKGPLLNFLTITAKDKRWEDPGKQLYNVNVEATSVALLALQLKDFDVPVVRWLNQOR	1260
DB	1201	RLKGPLLNFLTITAKDKRWEDPGKQLYNVNVEATSVALLALQLKDFDVPVVRWLNQOR	1260
QY	1261	YGGGYGSTQATPMVQALAQYQKADAPDHOELNLDVSLQPLSRSSKITTHRIHWSASLLR	1320
DB	1261	YGGGYGSTQATPMVQALAQYQKADAPDHOELNLDVSLQPLSRSSKITTHRIHWSASLLR	1320
QY	1321	SEETKNEGFTVTAEGKGOTLSVVTMWHAKAKDQTCNKFDLKVITIKPAPETERKPODA	1380
DB	1321	SEETKNEGFTVTAEGKGOTLSVVTMWHAKAKDQTCNKFDLKVITIKPAPETERKPODA	1380
QY	1381	KNTMILEICTRYRGDQDATMSILDISMMTGFAPTDDDLKQLANGVDVRIISKYELDKAFSD	1440
DB	1381	KNTMILEICTRYRGDQDATMSILDISMMTGFAPTDDDLKQLANGVDVRIISKYELDKAFSD	1440
QY	1441	RNTLIILYLDKVSHEDDCLAFKVHGVNVELLOPGAVKYAYYNLEESCTRFYHPEKEDG	1500
DB	1441	RNTLIILYLDKVSHEDDCLAFKVHGVNVELLOPGAVKYAYYNLEESCTRFYHPEKEDG	1500
QY	1501	KLNKLRDELRCABENCFIQKSDDKVLEERLDKACEFGVDYVYKTLRVKQLSNDPDE	1560
DB	1501	KLNKLRDELRCABENCFIQKSDDKVLEERLDKACEFGVDYVYKTLRVKQLSNDPDE	1560
QY	1561	YIMAIETQIKSGSDVQVQOQRTFISPIKCRALALEKXHYLMWGLSSDFWGEKPNLSY	1620
DB	1561	YIMAIETQIKSGSDVQVQOQRTFISPIKCRALALEKXHYLMWGLSSDFWGEKPNLSY	1620
QY	1621	IIGKDTWVHWPDEECQDEENKQCCQDLGAFTESMVFVFCPN 1663	
DB	1621	IIGKDTWVHWPDEECQDEENKQCCQDLGAFTESMVFVFCPN 1663	

RESULT 3  
 US-10-379-747-2  
 ; Sequence 2, Application US/10379747  
 ; Publication No. US20040023874A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Burgess, Catherine E.;  
 ; APPLICANT: Chant, John S.;

; APPLICANT: Chaudhuri, Amitabha ;  
 ; APPLICANT: Edinger, Shlomit R. ;  
 ; APPLICANT: Gangolli, Esha A. ;  
 ; APPLICANT: Gangolli, Uriel M. ;  
 ; APPLICANT: Miller, Charles E. ;  
 ; APPLICANT: Ooi, Chean Eng ;  
 ; APPLICANT: Ooi, Tatiana A. ;  
 ; APPLICANT: Patturajan, Meera ;  
 ; APPLICANT: Rastelli, Luca ;  
 ; APPLICANT: Rieger, Daniel K. ;  
 ; APPLICANT: Shimkets, Richard A. ;  
 ; APPLICANT: Zerhusen, Bryan D. ;  
 ; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
 ; FILE REFERENCE: 21402-568B  
 ; CURRENT APPLICATION NUMBER: US/10/379,747  
 ; CURRENT FILING DATE: 2003-03-05  
 ; PRIOR APPLICATION NUMBER: 60/365,034  
 ; PRIOR FILING DATE: 2002-03-15  
 ; PRIOR APPLICATION NUMBER: 60/366,420  
 ; PRIOR FILING DATE: 2002-03-21  
 ; PRIOR APPLICATION NUMBER: 60/365,477  
 ; PRIOR FILING DATE: 2002-03-19  
 ; NUMBER OF SEQ ID NOS: 45  
 ; SOFTWARE: Curaseq/blast version 0.1  
 ; SEQ ID NO 2  
 ; LENGTH: 1663  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-379-747-2

Query Match 99.9%; Score 8598; DB 15; Length 1663;  
 Best local Similarity 99.9%; Pred. No. 0;  
 Matches 1661; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MGPTSPSLLLLLLTHLPALGSPMYSITPNILRLESEETWLEAHDAQDVPVTVH	60
Db	1	MGPTSPSLLLLLLTHLPALGSPMYSITPNILRLESEETWLEAHDAQDVPVTVH	60
Qy	61	DFPGKLVLSSEKTVLTPATNMGNVTFTIPANREFKSEKGRNFVTVQATFGVQVEKV	120
Db	61	DFPGKLVLSSEKTVLTPATNMGNVTFTIPANREFKSEKGRNFVTVQATFGVQVEKV	120
Qy	121	VLYSLQGYLFQTDXTIYTPGSTVLYRIFTVNHKLLPVGRVTVMNIENPEGIPVKQDSL	180
Db	121	VLYSLQGYLFQTDXTIYTPGSTVLYRIFTVNHKLLPVGRVTVMNIENPEGIPVKQDSL	180
Qy	181	SSONQLGVLPDPLNDPELVNMGQWIRAYYENSPQGVFSTFEVKEVYLPSPFVIVEPTE	240
Db	181	SSONQLGVLPDPLNDPELVNMGQWIRAYYENSPQGVFSTFEVKEVYLPSPFVIVEPTE	240
Qy	241	KFYIYIYNEKGLVITITARELYGKVEGTAFVIFGIDQGEQRIISLPESLKRIPIEDGSGEV	300
Db	241	KFYIYIYNEKGLVITITARELYGKVEGTAFVIFGIDQGEQRIISLPESLKRIPIEDGSGEV	300
Qy	301	VLSRKVLGVQVQNPRAEDLVGSLVSATVILHSGSDMVQAEKRGIPVTSPIQIHFTKT	360
Db	301	VLSRKVLGVQVQNPRAEDLVGSLVSATVILHSGSDMVQAEKRGIPVTSPIQIHFTKT	360
Qy	361	PKYFKPGMPFDLMVFTNPDGSPAYRVPVAVQGEDVQSLTQGDGVAKLSINTHPSOKPL	420
Db	361	PKYFKPGMPFDLMVFTNPDGSPAYRVPVAVQGEDVQSLTQGDGVAKLSINTHPSOKPL	420
Qy	421	SITVTRTKKELSEAEQATMTQALPYSTVGNNSNYLHLSVLTRELPGETLNVNFLRMD	480
Db	421	SITVTRTKKELSEAEQATMTQALPYSTVGNNSNYLHLSVLTRELPGETLNVNFLRMD	480
Qy	481	RAHEAKIRYTYLIMNKGRLKAGQVREPGQDLVLPISLITTFIPSPFRIVAYTYLIGA	540
Db	481	RAHEAKIRYTYLIMNKGRLKAGQVREPGQDLVLPISLITTFIPSPFRIVAYTYLIGA	540
Qy	541	SGQREVVADSVVVDKSCVGLVVKSGSEDRQVPVQGMTLKIEGDHARVVLVAVDK	600
Db	541	SGQREVVADSVVVDKSCVGLVVKSGSEDRQVPVQGMTLKIEGDHARVVLVAVDK	600

Qy	601	GVFVLNKKNKLTQSKIMWVVEKADIGCTPGSGKDYAGVFSADAGLTFTSSGGQOTQRAEL	660
Db	601	GVFVLNKKNKLTQSKIMWVVEKADIGCTPGSGKDYAGVFSADAGLTFTSSGGQOTQRAEL	660
Qy	661	QCPOFAARRRRSVOLTTEKRMVKVGPKELRKCCEDGMRENPMRFSQRRTRFISLGAC	720
Db	661	QCPOFAARRRRSVOLTTEKRMVKVGPKELRKCCEDGMRENPMRFSQRRTRFISLGAC	720
Qy	721	KKVFLDCNCTITELRRHARASHLGLARSNLDEIIAENIVSRSEFFPESMLWNVEDLKE	780
Db	721	KKVFLDCNCTITELRRHARASHLGLARSNLDEIIAENIVSRSEFFPESMLWNVEDLKE	780
Qy	781	PPKNGISTKLMNIFPKDSITTWBILAVMSDDKKGICVADPFVTVQDFFIDRLRIPYSV	840
Db	781	PPKNGISTKLMNIFPKDSITTWBILAVMSDDKKGICVADPFVTVQDFFIDRLRIPYSV	840
Qy	841	RNEQVEIRAVLYNRYRQNELKRVVLLHNPAPFCSLATTKRHHQQITVTPPKSSLSVPYVI	900
Db	841	RNEQVEIRAVLYNRYRQNELKRVVLLHNPAPFCSLATTKRHHQQITVTPPKSSLSVPYVI	900
Qy	901	VPLTGLQVEVEVKAAYVHHFISGVVRKSLKVVPBGIRMNKTVAVRTLPERLGRGVQKE	960
Db	901	VPLTGLQVEVEVKAAYVHHFISGVVRKSLKVVPBGIRMNKTVAVRTLPERLGRGVQKE	960
Qy	961	DIPPADLSDOVPDTESETRILLQGTVPVQMTEDAVDAERLKHLLVTPSCGCGQNNMIGTP	1020
Db	961	DIPPADLSDOVPDTESETRILLQGTVPVQMTEDAVDAERLKHLLVTPSCGCGQNNMIGTP	1020
Qy	1021	TVIAVHVLDETEQWEKLEKRGALGLIKKGYTQQLAFQPSAPFAAFVKAEPSTWITA	1080
Db	1021	TVIAVHVLDETEQWEKLEKRGALGLIKKGYTQQLAFQPSAPFAAFVKAEPSTWITA	1080
Qy	1081	YVVKVFLAVNLIAIDSOVLGAVKWLILKQKPDGVFOEDAPVIFHQEMIGLRNNNEKD	1140
Db	1081	YVVKVFLAVNLIAIDSOVLGAVKWLILKQKPDGVFOEDAPVIFHQEMIGLRNNNEKD	1140
Qy	1141	MALTAFLVLSLOEAKDICEQVNSLPGSITKAGDFLEANYMLORSYTVAIAGYALAQM	1200
Db	1141	MALTAFLVLSLOEAKDICEQVNSLPGSITKAGDFLEANYMLORSYTVAIAGYALAQM	1200
Qy	1201	RLKGPLANKFLITAKDNRWEDPGKQLYNVEATSYALLALLQKDFDVPVVRWLNQOR	1260
Db	1201	RLKGPLANKFLITAKDNRWEDPGKQLYNVEATSYALLALLQKDFDVPVVRWLNQOR	1260
Qy	1261	YVGGYGSTQATPMVFOALAQYOKDAPDHQELNLDVSLQPSRSKITHRIHWESASLLR	1320
Db	1261	YVGGYGSTQATPMVFOALAQYOKDAPDHQELNLDVSLQPSRSKITHRIHWESASLLR	1320
Qy	1321	SEETKENEGETVTAEGKGGQGTLSVVTMYHAKAKDQTCNKFDLKVTIKPAPETERPQDA	1380
Db	1321	SEETKENEGETVTAEGKGGQGTLSVVTMYHAKAKDQTCNKFDLKVTIKPAPETERPQDA	1380
Qy	1381	KNTMLIETRYRGDQDATMSIIDIENMTGFAPDQDOLKOLANGVDVRYISKVELOKAFSD	1440
Db	1381	KNTMLIETRYRGDQDATMSIIDIENMTGFAPDQDOLKOLANGVDVRYISKVELOKAFSD	1440
Qy	1441	RNTLIIYLDKVSSEDDCLAFKHQVFNVELIQPGAVKYVAYYNLEESCTRFYHPEKEDG	1500
Db	1441	RNTLIIYLDKVSSEDDCLAFKHQVFNVELIQPGAVKYVAYYNLEESCTRFYHPEKEDG	1500
Qy	1501	KLNLKRDCLRCBAEENCFIQKSDDKVTLLEERLDKACEPGVDVYVYKTLVKQLNSDDE	1560
Db	1501	KLNLKRDCLRCBAEENCFIQKSDDKVTLLEERLDKACEPGVDVYVYKTLVKQLNSDDE	1560
Qy	1561	YIMAIEQITKSGDEVQVQQRFTFIPKICREALKLEKHYLMWGLSDFWGEKPNLSY	1620
Db	1561	YIMAIEQITKSGDEVQVQQRFTFIPKICREALKLEKHYLMWGLSDFWGEKPNLSY	1620
Qy	1621	IIGKDTWVHWPEDECCQDEENKQCDLGAFTTESMVVFGCPN	1663
Db	1621	IIGKDTWVHWPEDECCQDEENKQCDLGAFTTESMVVFGCPN	1663

```

US-10-174-333-41
; Sequence 41, Application US/10174333
; Publication No. US20040029220A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Zethusen, Bryan D.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Grosse, William M.
; APPLICANT: Szekeres, Edward S.
; APPLICANT: Alsobrook, John P.
; APPLICANT: Anderson, David W.
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Li, Li Mei
; APPLICANT: Zhong,
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-783 CIP1
; CURRENT APPLICATION NUMBER: US/10/174,333
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: 60/193,664
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/194,614
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: 60/195,063
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,066
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,067
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,068
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,069
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,070
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,510
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/219,855
; PRIOR FILING DATE: 2000-07-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: Curasequid version 0.1
; SEQ ID NO 41
; LENGTH: 1663
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-174-333-41

Query Match 99.9%; Score 8596; DB 15; Length 1663;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1661; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGPTSGPSLLLLTHLPLALGSPMYSIIIPNLRLESETWVLEAHDAGQDPVTVTVH 60
DB 1 MGPTSGPSLLLLTHLPLALGSPMYSIIIPNLRLESETWVLEAHDAGQDPVTVTVH 60
QY 61 DFPKGLVLSSEKIVLTPATNMGNTFTIPANREFKSEKGRNKFVTVOATFGTQWEKV 120

```

```

61 DFPKGLVLSSEKIVLTPATNMGNTFTIPANREFKSEKGRNKFVTVOATFGTQWEKV 120
121 VLVSQSGYLFIOCTDKTIYTPGSTVLYRIFTVNHKLLPVGRVTVNINENPGIPVKQDSL 180
121 VLVSQSGYLFIOCTDKTIYTPGSTVLYRIFTVNHKLLPVGRVTVNINENPGIPVKQDSL 180
181 SSQNLGVPLSWDIPELVNMGWKIRAYYENSPOQVFSTEFVEKYYVLPSPFVIEPTE 240
181 SSQNLGVPLSWDIPELVNMGWKIRAYYENSPOQVFSTEFVEKYYVLPSPFVIEPTE 240
241 KFYIYNEKGLVITITARELYGKKVEGTAFVIFGIDGEBORISLPESLKRPIEDGSGEV 300
241 KFYIYNEKGLVITITARELYGKKVEGTAFVIFGIDGEBORISLPESLKRPIEDGSGEV 300
301 VLSRKVLDDGVONPRAEDLVGKSLVYSATVILHSGSDMVQAPRSQIPVTPGYQHFTKT 360
301 VLSRKVLDDGVONPRAEDLVGKSLVYSATVILHSGSDMVQAPRSQIPVTPGYQHFTKT 360
361 PKYFKGMPFDMVFTNPDGSPAYRVPVAVOGEDEVQSLTQDGVAKLSINTHPSQKPL 420
361 PKYFKGMPFDMVFTNPDGSPAYRVPVAVOGEDEVQSLTQDGVAKLSINTHPSQKPL 420
421 SITVRTKQELSEAEQATMTQALPYSTVGNENNYLHLSVLRTELPRGETLVNPFLLMD 480
421 SITVRTKQELSEAEQATMTQALPYSTVGNENNYLHLSVLRTELPRGETLVNPFLLMD 480
481 RAHEAKIRYTYLIMNKGRLKAGQVRPQGDVLLVPLSIITDIPSPFRVAVYTLIGA 540
481 RAHEAKIRYTYLIMNKGRLKAGQVRPQGDVLLVPLSIITDIPSPFRVAVYTLIGA 540
541 SGQREVADSVVVDKSDCVGSLVWKSQSEDRQVPGQOMTKITEGHDGARVVLVAVDK 600
541 SGQREVADSVVVDKSDCVGSLVWKSQSEDRQVPGQOMTKITEGHDGARVVLVAVDK 600
601 GVFVLNKKNKLTOSKTDVVEKADIGCTPGSKDYGAVFSDAGLTFTSSSGQTAQRAEL 660
601 GVFVLNKKNKLTOSKTDVVEKADIGCTPGSKDYGAVFSDAGLTFTSSSGQTAQRAEL 660
661 QCPQPAARRRSVQLTEKMDKGVKPKELRKCCEGDMRENPMRPSQORTRFISLGAC 720
661 QCPQPAARRRSVQLTEKMDKGVKPKELRKCCEGDMRENPMRPSQORTRFISLGAC 720
721 KKVFLDCCNYITELRQHARASHGLARSLNLDIIAENIVSRSEFFESWLNVEDLKE 780
721 KKVFLDCCNYITELRQHARASHGLARSLNLDIIAENIVSRSEFFESWLNVEDLKE 780
781 PPKNGISTKLMIIFLKDSITTWELLAVSMDSKKGICVADPPEVTVMQDFFIDLRLPYSV 840
781 PPKNGISTKLMIIFLKDSITTWELLAVSMDSKKGICVADPPEVTVMQDFFIDLRLPYSV 840
841 RNEQVEIRAVLYNRYNQELKVRVELLHNPFCSLATTKRRHQQTITIPPKSSLSVPYVI 900
841 RNEQVEIRAVLYNRYNQELKVRVELLHNPFCSLATTKRRHQQTITIPPKSSLSVPYVI 900
901 VPLKTGLQEVKAAVYHHFISDGVKSLKVVPEGIRNKNVAVRTLDPKREGVQKE 960
901 VPLKTGLQEVKAAVYHHFISDGVKSLKVVPEGIRNKNVAVRTLDPKREGVQKE 960
961 DIPPADLSQVDPDTESETRILLQGTPTVAQMTEDAVABRLKHLIIVTPSGCGEONMIGTP 1020
961 DIPPADLSQVDPDTESETRILLQGTPTVAQMTEDAVABRLKHLIIVTPSGCGEONMIGTP 1020
1021 TVTAVHYLDETEQWEKFGLEKQGALELIKGYTQQLAFRQPSAFAFVVKRAESTWTLTA 1080
1021 TVTAVHYLDETEQWEKFGLEKQGALELIKGYTQQLAFRQPSAFAFVVKRAESTWTLTA 1080
1081 YVVKVPSLAVNLIAIDSVLCGAVKWLILEKQKPGVFEQEDAPVTHQEMIGGLRNNNEKD 1140
1081 YVVKVPSLAVNLIAIDSVLCGAVKWLILEKQKPGVFEQEDAPVTHQEMIGGLRNNNEKD 1140
1141 MALTAFVLISQBAKDICEEQVNSLPGSITKAGDFLEANYMNLQRSYTVIAGYALAOQG 1200
1141 MALTAFVLISQBAKDICEEQVNSLPGSITKAGDFLEANYMNLQRSYTVIAGYALAOQG 1200

```

1201 RLKGPLNKLFTAKDKNRPEDGKOLYNVEATSVALLALLQLKDFDFVPPVWVWLNQOR 1260  
1201 RLKGPLNKLFTAKDKNRPEDGKOLYNVEATSVALLALLQLKDFDFVPPVWVWLNQOR 1260  
1261 YGCGYGSQTATPMVQALAQKQAPDHPQELNLDVSLQPSRSKITHRIHWESASLLR 1320  
1261 YGCGYGSQTATPMVQALAQKQAPDHPQELNLDVSLQPSRSKITHRIHWESASLLR 1320  
1321 SEETKNEGFTVTAEGKGQGTISVVMYHAKAKDQLTCKNPKLVYTIKPAPETEKRPQDA 1380  
1321 SEETKNEGFTVTAEGKGQGTISVVMYHAKAKDQLTCKNPKLVYTIKPAPETEKRPQDA 1380  
1381 KNTMLIEICTRYRGQDQATMSILDISMGTGAPDPTDDLKQLANGVDVRYISKYELDKAFSD 1440  
1381 KNTMLIEICTRYRGQDQATMSILDISMGTGAPDPTDDLKQLANGVDVRYISKYELDKAFSD 1440  
1441 RNTLIYLDKVSSEDDCLAFKHQYFNVVELIOPGAVKYVAYYNLEESCTRFYHPEKEDG 1500  
1441 RNTLIYLDKVSSEDDCLAFKHQYFNVVELIOPGAVKYVAYYNLEESCTRFYHPEKEDG 1500  
1501 KLNKLCRDELCRAENCFIQSDDKVTLLEERLDKACBPVDVYVYKTLVKVQLSNDFDE 1560  
1501 KLNKLCRDELCRAENCFIQSDDKVTLLEERLDKACBPVDVYVYKTLVKVQLSNDFDE 1560  
1561 YIMAEOTIKSGSDEVOVGQORTFISPIKCREALKLEEKHYLMWGLSSDFWGEKPNLSY 1620  
1561 YIMAEOTIKSGSDEVOVGQORTFISPIKCREALKLEEKHYLMWGLSSDFWGEKPNLSY 1620  
1621 IIGKOTVWEHPDEDCQDEENKQKQCDLIGATESVWVFGCPN 1663  
1621 IIGKOTVWEHPDEDCQDEENKQKQCDLIGATESVWVFGCPN 1663

RESULT 5  
US-09-842-758-43  
; Sequence 43, Application US/09842758  
; Publication No. US20030083244A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A. M.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Malvankar, Uriel M.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Zernusen, Bryan D.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Majumder, Rumud  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Rastelli, Luca  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Taupier, Raymond J.  
; APPLICANT: Grosse, William M.  
; APPLICANT: Edward, Szekeres S.  
; APPLICANT: Alagbrook II, John P.  
; TITLE OF INVENTION: No. US20030083244A1e1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-783  
; CURRENT APPLICATION NUMBER: US/09/842, 758  
; CURRENT FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/200,158  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: 60/200,613  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,780  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/201,006  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,007

PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,236  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,238  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,186  
PRIOR FILING DATE: 2000-05-02  
PRIOR APPLICATION NUMBER: 60/201,474  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/201,508  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/220,591  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: 60/232,678  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 60/263,217  
PRIOR FILING DATE: 2001-01-22  
PRIOR APPLICATION NUMBER: 60/265,160  
PRIOR FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 113  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 43  
LENGTH: 1663  
TYPE: PR1  
ORGANISM: Mus musculus  
US-09-842-758-43

Query Match 78.4%; Score 6752; DB 10; Length 1663;  
Best Local Similarity 77.1%; Pred. No. 0;  
Matches 1285; Conservative 178; Mismatches 197; Indels 6; Gaps 5;

QY 1 MGPTSGPSL--LILLTLHLPLALGSPMYSIITNIIELSEETMVLHAHQDQVPVTVT 58  
DB 1 MGPASGQLLVLLILLASSPLALGIPMYSIITNIIELSEETMVLHAHQDQVPVTVT 60  
QY 59 VHPFGKLVLSSEKTVLTATNMGVNTFTIIPANREFSEKGRNKFVTVQATFGTVVE 118  
DB 61 VQDFL-KRQVLTSEKTVLTGASHLRSVSIKIPASKEFNSDKGKHYVTVANFETVVE 119  
QY 119 KVLVLSQSGYLFITDKTITPGSTVLYRIFTNNHKLPLVGTNNVNIENEGIPVKOD 178  
DB 120 KAVVVSFQSGYLFITDKTITPGSTVLYRIFTNNHKLPLVGTNNVNIENEGIPVKOD 179  
QY 179 SLSSONQLVPLSDWIPELVNMGMOWIRAYVNSPOOVFSTEFVEKVVLPSEFVVEP 238  
DB 180 ILSSNNQHGILPLSWNIPELVNMGMOWIRAFYEHAKQIFSAEFVEKVVLPSEFVVEP 239  
QY 239 TEKYYIYNEKGLVTTIARFLYKQKVEGTAFVIFGIQDGEQRISELSKRIPIEDGSG 298  
DB 240 TETFYIIDDPNGLEVSIIAKFLYKGNVDGTAFVIFGVQDGGKKISLAHSLTRVIEDGV 299  
QY 299 EVLRSKVLDDGVNPRABDLVGLSVSATVILHSGSDMVQAERSGIPVITSPYQIHT 358  
DB 300 DAVLTRKVLMEGVTPSNADALVGLSVSATVILHSGSDMVQAERSGIPVITSPYQIHT 359  
QY 359 KTKPKFKPMFPDLWVFTVTPDGSPPAYRVAVQGEDTVQSLTQGGVAKLSINTHPSOK 418  
DB 360 KTKPKFKPMFPDLWVFTVTPDGSPPAYRVAVQGEDTVQSLTQGGVAKLSINTHPSOK 418  
QY 419 PLSTIVTRTKKQELSEASQATRTMQLPSTVGNNSNYHLSTRLRPGETINNVNLLR 478  
DB 419 PLSTIVTRTKKQELSEASQATRTMQLPSTVGNNSNYHLSTRLRPGETINNVNLLR 478  
QY 479 MDRAHEAKIRYTYLNNKRLKAGQVREPODLVPLSTITTDIFISFLVAYTYLI 538  
DB 479 TDPGHEAKIRYTYLNNKRLKAGQVREPODLVPLSTITTDIFISFLVAYTYLI 538  
QY 539 GASGQREVVADSVWVDVKDCVGLSVKSGSDRQVPCQQTALXIEGHGAVRVLVAV 598  
DB 539 GASGQREVVADSVWVDVKDCVGLSVKSGSDRQVPCQQTALXIEGHGAVRVLVAV 598  
QY 599 DKGVFLNNKKLQSKINWVWEKADIGCTPGSGKDYAGVFSAGLITFTSSSQQAORA 658  
DB 599 DKGVFLNNKKLQSKINWVWEKADIGCTPGSGKDYAGVFSAGLITFTSSSQQAORA 658





Db 120 KAMVBSFQSGFIQIDKTIYTPGSTVLYRIFTVDNNLLPVGKTVWLLIETPDGIPVKRD 179  
Qy 179 SLSSQNLGVLPLSWDIPBLVNMGNKIRAYIENSPOQVFTSEFEVKEYVLPSPSEVIVEP 238  
Db 180 ILSNNQHGLPILSWNIPBLVNMGNKIRAFYEHAPKQIFABSEFEVKEYVLPSPSEVIVEP 239  
Qy 239 TEFXYIYNEKGLLEVITAFRLYGLKVEGTAFVIFIGIQDGEORSLPESIKRIPIEDGSG 298  
Db 240 TEFXYIIDDPNGLEVISIAKFLYGNVDGTAFFVFGVDGDKISLAHSLTRVVEDVG 299  
Qy 299 EVULSRKVLGDVONPAEDINGKSLYSATVILHSGSDMVQABRSIGIPVTSYQIHFT 358  
Db 300 DAVLTRKVLMEGVRPNADALVGLSYSVTVILHSGSDMVQABRSIGIPVTSYQIHFT 359  
Qy 359 KTKYFKPMPFDMVFTVNPDSGPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSOK 418  
Db 360 KTKYFKPMPFDMVFTVNPDSGPASKVLVVTQGSN-AKALTQDDGVAKLSINTHPSRQ 418  
Qy 419 PLSITVTRKQELSEABQATRWQALPYSTVGNSSNNYLHLSVLRTELAPGETLNVNLLR 478  
Db 419 PLTITVTRKQELSEABQATRWQALPYSTVGNSSNNYLHLSVLRTELAPGETLNVNLLR 478  
Qy 479 MDRAEAKIRYVYTLIMNKGRLKAGQVREPGQDLVPLSLITDTIPSPRLVAYYTLI 538  
Db 479 TDPGHEAKIRYVYTLIMNKGRLKAGQVREPGQDLVPLSLITDTIPSPRLVAYYTLI 538  
Qy 539 GASQREVADSVWVDVKDSCVGLSVKSGQSEDEQVPVPGQOMLKTEGDHGAHVILVAV 598  
Db 539 GASQREVADSVWVDVKDSCVGLSVKSGQSEDEQVPVPGQOMLKTEGDHGAHVILVAV 597  
Qy 599 DKGVFVNLKKNKLTQSKIWDVVEKADIGCTPGSGKDYAGVFSADAGLTFTSSSGQQTQARA 658  
Db 599 DKGVFVNLKKNKLTQSKIWDVVEKADIGCTPGSGKDYAGVFSADAGLTFTSSSGQQTQARA 657  
Qy 659 ELQCPQAPARRRSVOLTERRKVDKGY-PRELKCCEDGMRENMFSCORRFPFSLG 717  
Db 659 ELQCPQAPARRRSVOLTERRKVDKGY-PRELKCCEDGMRENMFSCORRFPFSLG 717  
Qy 718 EACKVFLDCNYITELRQHARASHGLGARSNDELIIAENIVTSRSEPPESLMNVED 777  
Db 718 ENCIKAFIDCNHITKLRQHRDHDVGLGARSELEEDIIPEEDIISRHSPQSLWLTIE 777  
Qy 778 LKSPKNGISTKLNNIFLQSTITWEILAVSNDKKGICVADPPEVTMQDPFDILRLPY 837  
Db 778 LKSPKNGISTKLNNIFLQSTITWEILAVSNDKKGICVADPPEVTMQDPFDILRLPY 837  
Qy 838 SVVRNEQVEIRAVLYNRYQNELKVRVELLHNPAFCSLATTKRRHQQTITIPKSSLSVP 897  
Db 838 SVVRNEQVEIRAVLYNRYQNELKVRVELLHNPAFCSLATTKRRHQQTITIPKSSLSVP 897  
Qy 898 YVAVBLKTGLQVEVKAAYHYHPI SDGYRKSIVKVPGEIRMNKTVAVRTLDPELREGV 957  
Db 898 YVAVBLKTGLQVEVKAAYHYHPI SDGYRKSIVKVPGEIRMNKTVAVRTLDPELREGV 957  
Qy 958 QKEDIPADISQVPTSETRIILQGTTPVAGQMTEDAVDAERLKHIVTPSGCGEQNMIG 1017  
Db 958 QKEDIPADISQVPTSETRIILQGTTPVAGQMTEDAVDAERLKHIVTPSGCGEQNMIG 1017  
Qy 1018 MTETVAVHVLDETEQWEGFLEKCGALELKKGYTOQLAFROPSSAPAFVREASTW 1077  
Db 1018 MTETVAVHVLDETEQWEGFLEKCGALELKKGYTOQLAFROPSSAPAFVREASTW 1077  
Qy 1078 LTAYVVKVSLAVNLIAIDSVLCGAVKWLILEKQKPGVFDQEDAPVTHQEMIGGLRNN 1137  
Db 1078 LTAYVVKVSLAVNLIAIDSVLCGAVKWLILEKQKPGVFDQEDAPVTHQEMIGGLRNN 1137  
Qy 1138 EKDMALTAFVLISLOAKDICEQVNSLPGSTTKAGDFLEANYMQLORSYVATAGYALA 1197  
Db 1138 EKDMALTAFVLISLOAKDICEQVNSLPGSTTKAGDFLEANYMQLORSYVATAGYALA 1197  
Qy 1198 QMGLKGPPLNKLTAKDKNRWEDPGKOLYNVEATSYALLALILQDKDFDVPVVRWLN 1257  
Db 1198 QMGLKGPPLNKLTAKDKNRWEDPGKOLYNVEATSYALLALILQDKDFDVPVVRWLN 1257

Db 1198 LMNKLSEPYLKPNTAKDRNWEEDQQLYNVEATSYALLALILQKDFSDVPVVRWLN 1257  
Qy 1258 EORVYGGYGSTQATFMVFOALAQYOKDAPDHOELNDVSLQPSRSSKITHRIHWESAS 1317  
Db 1258 EORVYGGYGSTQATFMVFOALAQYQIDVDPDKDNDVSVFHPSSRSATITRLLWENG 1317  
Qy 1318 LLRESEETKNEGFTVTAEGRGQGTLSVVTMYHAKAKDQLTQNKFDLKVYTIKPAPETERP 1377  
Db 1318 LLRESEETKNEGFTVTAEGRGQGTLSVVTMYHAKAKDQLTQNKFDLKVYTIKPAPETERP 1377  
Qy 1378 QDANKWILECTEYRGDQDATMSIILDISMTGTFAPDQDOLKOLANGVDVYISKVELDKA 1437  
Db 1378 QDANKWILECTEYRGDQDATMSIILDISMTGTFAPDQDOLKOLANGVDVYISKVELDKA 1437  
Qy 1438 FSDRNTLIIYLDKVSHEDECLAFKVHQYFNVELIQGAVKYVAYYNNLEESCTRFYHPEK 1497  
Db 1438 FSDRNTLIIYLDKVSHEDECLAFKVHQYFNVELIQGAVKYVAYYNNLEESCTRFYHPEK 1497  
Qy 1498 EDGKLNKLCRDELRCRAEENCFIQKSDDKVTLLEERLDKACBQVDVYVYKTRLVQLSND 1557  
Db 1498 EDGKLNKLCRDELRCRAEENCFIQKSDDKVTLLEERLDKACBQVDVYVYKTRLVQLSND 1557  
Qy 1558 FDEYMAIEQCTIKSGSDEVQVQGTFTFIPKICREALKLEEKHYLMGLSSDFWGEKPN 1617  
Db 1558 FDEYMAIEQCTIKSGSDEVQVQGTFTFIPKICREALKLEEKHYLMGLSSDFWGEKPN 1617  
Qy 1618 LSYIIGKDTWVHWPEDECOEDENOKOQODLGAFTESMVVFGCPN 1663  
Db 1618 LSYIIGKDTWVHWPEDECOEDENOKOQODLGAFTESMVVFGCPN 1663

## RESULT 7

US-09-842-758-42  
Sequence 42, Application US/09842758  
Publication No. US20030083244A1  
GENERAL INFORMATION:  
APPLICANT: Vernet, Corine A. M.  
APPLICANT: Fernandes, Ema R  
APPLICANT: Gerlach, Valerie  
APPLICANT: Shimkets, Richard A  
APPLICANT: Malyankar, Uriel M  
APPLICANT: Boldog, Ferenc L  
APPLICANT: Zethusen, Bryan D  
APPLICANT: Spytek, Kimberly A  
APPLICANT: Majumder, Kumud  
APPLICANT: Tchernev, Velizar T  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Patturajan, Meera  
APPLICANT: Burgess, Catherine E  
APPLICANT: Gangolli, Esha A  
APPLICANT: Smithson, Glennnda  
APPLICANT: Rastelli, Luca  
APPLICANT: MacDougall, John R  
APPLICANT: Taupier, Raymond J  
APPLICANT: Grosse, William M  
APPLICANT: Edward, Szekeres S  
APPLICANT: Alsobrook II, John P  
TITLE OF INVENTION: No. US20030083244A1e1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-783  
CURRENT APPLICATION NUMBER: US/09/842,758  
CURRENT FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/200,158  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/200,613  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,780  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/201,006  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,007  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,236  
PRIOR FILING DATE: 2000-05-01

/	PRIOR APPLICATION NUMBER:	60/201,238
/	PRIOR FILING DATE:	2000-05-01
/	PRIOR APPLICATION NUMBER:	69/201,186
/	PRIOR FILING DATE:	2000-05-02
/	PRIOR APPLICATION NUMBER:	60/201,474
/	PRIOR FILING DATE:	2000-05-03
/	PRIOR APPLICATION NUMBER:	60/201,508
/	PRIOR FILING DATE:	2000-05-03
/	PRIOR APPLICATION NUMBER:	60/220,591
/	PRIOR FILING DATE:	2000-07-25
/	PRIOR APPLICATION NUMBER:	60/232,678
/	PRIOR FILING DATE:	2000-09-15
/	PRIOR APPLICATION NUMBER:	60/263,217
/	PRIOR FILING DATE:	2001-01-22
/	PRIOR APPLICATION NUMBER:	60/265,160
/	PRIOR FILING DATE:	2001-01-30
/	NUMBER OF SEQ ID NOS:	113
/	SOFTWARE:	Patentin Ver. 2.1
/	SEQ ID NO 42	:
/	LENGTH:	1661
/	TYPE:	PRT
/	ORGANISM:	Sus scrofa
/	US-09-842-758-42	:
Query Match            78.1%   Score 6724; DB 10; Length 1661;		
Best Local Similarity   76.5%   Pred No 0;		
Matches 1272; Conservative 189; Mismatches 200; Indels 2; Gaps 2;		
QY	1	MGPTSGPSLLLLLTHPLALGSPMYSITTPNILESEETWVLEBAHQDGPVTVTVH 60
Ddb	1	MGSTSGPRLLLLTSLFALGDPIYTIITPNVLRLESEMVLSEAHEGGQDIRSVTVH 60
QY	61	DFPGKLVASEKTIVLPATNHMGNTFTIPANKREPKEBGNKNFTVQAATFGTQVVEKV 120
Ddb	61	DFPAKRQVLSAETTNNANNYLSTVNKIIPASKEFKSEKG-HKFTVTQALPGNVQVEKV 119
QY	121	VLVLSQGYLFIQDTXTIYPGTSVIYRIFTVNHKLLPVGRVTMVNIENPSGIPVKQDSL 180
Ddb	120	VLVLSQGYLFIQDTKYITPYGTSVIYRIFTVDHKLLPVGQTIVVITETPEGIDIKRDSL 179
QY	181	SSQNQLGVPLSDWDIPELVNMNGOWKIRAYENSPOOVSTEBEVEKEYVLPSEFIVRPTE 240
Ddb	180	SSHNQFGILALSNNIPELVNMNGOWKIRAHVEDAPQOVSAEFVEKEYVLPSEFIVRPTE 239
QY	241	KFYIYNKGLVTTIARFLYKKVBGTAFVIFGIODGEQRISLPESUKRPIEDGSGEV 300
Ddb	240	KFYIIDDPNGLTWNIAARFLYGESVGTAFVIFGQDGQRISLSQSUTRVPFIIDGTGEA 299
QY	301	VLRSKVLDDGVONPRABDLVGKLYVSATVILHSGSDMWCAERSGIPVTSPYQHFTKT 360
Ddb	300	TLSQGVLINGVHYSSVNDLVGKIYVSVTILNSGSDMWAEARTGIPVTSPTYQHFTKT 359
QY	361	KPKFGMPFDLVFVNTNPDSFPAYRVPAVOGEDTVQSLTGQDVAKLSINTHPSKPL 420
Ddb	360	KPKFKAMPFDLVYVNTNPDSFPARI PVVTE-DFKVRSLTDQEDGVAKLSINTPDNRSL 418
QY	421	SITVTRKKQELSRAEQATRMQALPYSTVGENSNYLIHLSELTELPCGETLNVPFLRMD 480
Ddb	419	PIIVTEKDGI PAARQASKTMHVL PYNTQENSNYLIHLSPVELKPGENLVNFHLRTD 478
QY	481	RAHBAKRYTYTLIMNKGRLLKAGROVRPBGDDLVLPLSITTDFIPSRFLVAYTII GA 540
Ddb	479	PGYQDKIRYFTYLI MNKGKLLKVGROFRSGQVVVPLPTIITDFIPSRFLVAYTII AA 538
QY	541	SGOREVWADS VVWDVKDS CVGSLVXKSQSEDRQPVPQQMTLKIEGHGARGVVLVAVDK 600
Ddb	539	NGOREVWADS VVWDVKDS CGTVLVKGGGQDKQHRPQQMTLEIQGERGARVLVAVDK 598
QY	601	GVFLVKNKNKLTKSQKHUVVEKADICTGPGSGKDVAGVPSDAGLFTTSSSCQOAA RAE L 660
Ddb	599	GVFLVKNKNKLTKORRI MWVVEKADI CTGPGSGKFAGVTFDTAGLAFKFSKGLQTPQ S AD L 658
QY	661	QCQOPFAARRRRSVGLTEKRNMDKVGYKPKELRKCCEDGMRENPMPSCORRTFRFISLGEAC 720

APPLICANT: Vernet, Corine A.M.  
 APPLICANT: Fernandes, Elma R.  
 APPLICANT: Gerlach, Valerie  
 APPLICANT: Malyankar, Uriel M.  
 APPLICANT: Boldog, Ferenc L.  
 APPLICANT: Zerhusen, Bryan D.  
 APPLICANT: Seytek, Kimberly A.  
 APPLICANT: Mejunier, Kumud  
 APPLICANT: Tchernev, Velizar T.  
 APPLICANT: Patturajan, Meera  
 APPLICANT: Burgess, Catherine E.  
 APPLICANT: Gangolli, Bsha A.  
 APPLICANT: Smithson, Glenda  
 APPLICANT: Rastelli, Luca  
 APPLICANT: MacDougall, John R.  
 APPLICANT: Taupier, Raymond J.  
 APPLICANT: Grosse, William M.  
 APPLICANT: Szekeres, Edward S.  
 APPLICANT: Alsobrook, John P.  
 APPLICANT: Anderson, David W.  
 APPLICANT: Guo, Xiaojia (Sasha)  
 APPLICANT: Li, Li  
 APPLICANT: Zhong, Mei  
 TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
 FILE REFERENCE: 15966-783 CIP1  
 CURRENT APPLICATION NUMBER: US/10/174,333  
 CURRENT FILING DATE: 2002-06-18  
 PRIOR APPLICATION NUMBER: 60/193,664  
 PRIOR FILING DATE: 2000-03-31  
 PRIOR APPLICATION NUMBER: 60/194,614  
 PRIOR FILING DATE: 2000-04-05  
 PRIOR APPLICATION NUMBER: 60/195,063  
 PRIOR FILING DATE: 2000-04-06  
 PRIOR APPLICATION NUMBER: 60/195,066  
 PRIOR FILING DATE: 2000-04-06  
 PRIOR APPLICATION NUMBER: 60/195,067  
 PRIOR FILING DATE: 2000-04-06  
 PRIOR APPLICATION NUMBER: 60/195,068  
 PRIOR FILING DATE: 2000-04-06  
 PRIOR APPLICATION NUMBER: 60/195,069  
 PRIOR FILING DATE: 2000-04-06  
 PRIOR APPLICATION NUMBER: 60/195,070  
 PRIOR FILING DATE: 2000-04-06  
 PRIOR APPLICATION NUMBER: 60/195,510  
 PRIOR FILING DATE: 2000-04-06  
 PRIOR APPLICATION NUMBER: 60/219,855  
 PRIOR FILING DATE: 2000-07-21  
 Remaining Prior Application data removed - See File Wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 186  
 SOFTWARE: CuroSeqlist version 0.1  
 SEQ ID NO 42  
 LENGTH: 1661  
 TYPE: PRT  
 ORGANISM: Sus scrofa  
 US-10-174-333-42

Query Match 78.1%; Score 6724; DB 15; Length 1661;  
 Best Local Similarity 76.3%; Pred. No. 0;  
 Matches 1272; Conservative 189; Mismatches 200; Indels 2; Gaps 2;

QY 1 MGSTGSPSLLLTLLPLALGSPMYSIITPNILRLESEETWVLEAHDAQGDVPTVTVH 60  
 DB 1 MGSTGSPRLLLLSLPLALGDPVITITPNVLRLESEETWVLEAHFGQGDVSVTVH 60  
 QY 61 DFPQKLVLSSEKTLTPATNMGNTFTIPANRFFKSEKGNKNTVTVQATFGQVEKV 120  
 DB 61 DFPKRVLSSESTTLNANNYSITVNIIPAKSEKSEK-KHEVTQALFGNVQVEKV 119  
 QY 121 VLVSQSGYLFIQTDKTIYTPGSGTVLYRIFTVNHKLLPVGRVTVMNINENPSGIPVKQDSL 180  
 DB 120 VLVSQSGYLFIQTDKTIYTPGSGTVLYRIFTVDHKLPLVQGIIVTITPTSGIDIKRDSL 179

181 SSQNLQGLVPLSWDIPELVNMGMKIRAYYENSPQVSTEFPEVKYVLPSPSEVIVEPTE 240  
 DB 180 SSHNQFGLIALSNIPELVNMGMKIRAHYEDAPQVFSAEFEVKYVLPSPSEVIVEPSE 239  
 QY 241 KFYIYNEKGLVITITAFYLGKVEGTAFVIGIQDGEORISLPSLSKRIPIEGSGSV 300  
 DB 240 KFYIIDDENGTLVNIARFLYGESVDGTAFFVIGVGDGDRISLSQSLRFLIDGTGBA 299  
 QY 301 VLSRKVLDDGVNPRADLYGKSLYVSATVILHSGSDMVQABRSGLPIVTSYQIHFKT 360  
 DB 300 TLSQGVLLNGVHYSSVNDLVGKSIYVSIVILNSGSDMVEAERTGPIVTSYQIHFKT 359  
 QY 361 PKYFKPGMPFDMVFTNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHSPKPL 420  
 DB 360 PKFKFAPMPFDMVYVTPDGSFARHIPVTE-DFKVRSLTQEDGVAKLSINTPDNRSL 418  
 QY 421 SITVTKKQELSEABQATRTMQALPYSTVGNNSNYLHSLVRLTELPFGTLNVELLRMD 480  
 DB 419 FITVTERXDGIPAAARQASKTMHVLVYNTQNSKNYHLSLPRVELKPGENLVNPHLRD 478  
 QY 481 RAHEAKIRYTYLIMNKGRLLKAGQVRBPGQDLVPLSLITITDFIPSPRLVAYYTLGA 540  
 DB 479 PGYQDKIRYTYLIMNKGLLVGRQPRSGQVWVPLPLTITITDFIPSPRLVAYYTLGA 538  
 QY 541 SCOREVADSVWVVKSCVGSVWVKSQSEDQPVPGQOMTLKIEGDHARVWLVAVDK 600  
 DB 539 NSQREVADSVWVVDKDCVGTLLVKGSGKQKHQHPGQOMTLKIEQGERGARVGLVAVDK 598  
 QY 601 GVFLNKKNKLTQSKIWDVVEKADIGCTPGSGKDYAGVFSADGLTFTSSGQQTQARBL 660  
 DB 599 GVFLNKKNKLTORRINDVVEKADIGCTPGSGKDFAGVFTDAGLAFKSKGLQTPGRADL 658  
 QY 661 QCPQPAARRRSVQITKRMQKVGKPKELKRCCEGMRENPMRFSQORTAFISLGEAC 720  
 DB 659 ECPKPAARRRSVQIMKRMQKVGKPKELKRCCEGMRENPMRFSQORTAFISLGEAC 718  
 QY 721 KKVFLDCCNITELRRQHARASHGLARSNLDDEITAEENIVSRSEFPESWLNVEDLKE 780  
 DB 719 VKAFDCEYIAKLQKHSRKNKPLGLARSLDEIPEEDIISRSQFPESWLTVEEPE 778  
 QY 781 PPKNGISTKLNIFLKDSITTWELAVMSDKGICVADPFVTVVMDFFIDLRIPYSV 840  
 DB 779 PDKNGISTKLNIFLKDSITTWELAVMSDKGICVADPFVTVVMDFFIDLRIPYSV 838  
 QY 841 RNEQVEIRAVLYNQVQELKVRVELLHNPAPCSLATAKKGHQTLTVPAKSSVPVPII 900  
 DB 839 RNEQVEIRAVLYNQVQELKVRVELLHNPAPCSLATAKKGHQTLTVPAKSSVPVPII 898  
 QY 901 VPLKTGLQEVVKAHVHFIISDQVRSKLVVPEGIMKNTVAVRILDPERLREGVOKE 960  
 DB 899 VELKTGLQEVVKAHVHFIISDQVRSKLVVPEGIMKNTVAVRILDPERLREGVOKE 958  
 QY 961 DIPPADLSDQVDPDTESETRILLOQTPVAQMTEDAVERLKLHIVTFSGCGEONMGTP 1020  
 DB 959 EIPPADLSDQVDPDTESETRILLOQTPVAQMTEDAVERLKLHIVTFSGCGEONMGTP 1018  
 QY 1021 TVIAVHYLDTEQWEKFGLEKQCALIKKGYTOOLAPQPSAAFAKVPAPSWLTA 1080  
 DB 1019 TVIAVHYLDTEQWEKFGLEKQCALIKKGYTOOLAPQPSAAFAKVPAPSWLTA 1078  
 QY 1081 VYVVFSLAVNLIAIDSVQLCGAVKWLILEKQKPDGVQFQEDAFVHQEMIGGLNNNEKD 1140  
 DB 1079 VYVVFSLAVNLIAIDSVQLCGAVKWLILEKQKPDGVQFQEDAFVHQEMIGGLNNNEKD 1138  
 QY 1141 MALTAFLVLSIOEAKDICEQVNSLPGSITKAGDFLEANYMLQSVTVVAIAGYALAOXG 1200  
 DB 1139 VSLTAFLVLSIOEAKDICEQVNSLPGSITKAGDFLEANYMLQSVTVVAIAGYALAOXG 1198  
 QY 1201 RLKGPLLNKLTFTAKRNRWEDPGQLYNVEATSYALLLQKDFDPPVPPVVRWLNNEOR 1260  
 DB 1199 KLDEPFLNKLSTAKRNRWEDPGQLYNVEATSYALLLQKDFDPPVPPVVRWLNNEOR 1258  
 QY 1261 YVGGYGVSTQATFWVQALAQYQKADPHOELNLDVSLQPSRSSKITHIHWEASALLR 1320

```

Db 1259 YGCGYGTQATWVFOALAQYQKQVDPDHKDLNLDVSIHLPSAPVRHRIILWESALLR 1318
QY 1321 SEETKEGEGTVTAEGKGQOTLSVWYHYAKAQOJTCNKFDLKVTKPAPETERPQQA 1380
Db 1319 SEETKEGEGTVTAEGKGQOTLSVWYHYAKAQOJTCNKFDLKVTKPAPETERPQQA 1378
QY 1381 KNTMILICIRYRQDQATMSIILDISMTGFAPDQDOLQLANGVDVYISKYELDKAFSD 1440
Db 1379 KSSMVLDCIRYLNQDQATMSIILDISMTGFSPDTEDLKLSFGVDVYISKYELNALS 1438
QY 1441 NNTLIIYLDKUSHEDCLAPKHQVFNVELIOPGAVKYVAYYNLEESCTRYHPEKEDG 1500
Db 1439 KNTLIIYLDKUSHEDCLAPKHQVFNVELIOPGAVKYVAYYNLEESCTRYHPEKEDG 1498
QY 1501 KNLKCLDELCRAEENCFQKSDCKVTLEERLDKACEGVDYVYKTRLVKQLSNDPDE 1560
Db 1499 MLNKLCHKRCRAEENCFVHDEEVEVTLDDRERACEGVDYVYKTRLLKELSDPDD 1558
QY 1561 YIMAEIOTIISGSDDEVQVQQRFTFISPIKCBALKLEBKHYLMWGLSDPFGKPNLSY 1620
Db 1559 YIMAEIOTIISGSDDEVQVQQRFTFISPIKCBALKLEBKHYLMWGLSDPFGKPNLSY 1618
QY 1621 IIGKDTWVHPWDEDEQDQENKQKQODLGAFTESMVVFGCPN 1663
Db 1619 IIGKDTWVLPDGDVQDQENKQKQODLANFSENWVFGCPN 1661

```

## RESULT 9

```

US-09-925-442-2
; Sequence 2, Application US/09925442
; Patent No. US20020103346A1
; GENERAL INFORMATION:
; APPLICANT: VOGEL, CARL-WILHELM
; BREDEHORST, REINHORST
; KOCK, MICHAEL
; FRITZINGER, DAVID
; TITLE OF INVENTION: RECOMBINANT PROCVF
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/925,442
; FILING DATE: 10-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/017,947
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1126-0107-0X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1642 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-925-442-2

```

```

Query Match 49.1%; Score 4224; DB 9; Length 1642;
Best Local Similarity 49.6%; Pred. No. 0;
Matches 824; Conservative 324; Mismatches 478; Indels 36; Gaps 20;

QY 9 LLLLLLTHLPALGSPMYIIITPILRLSEETMVLAEHDAQDQDPVVTVHDPFGKLV 68
Db 8 LVAALLIGPSSSHGALYTLITPAVLRTDTTEQILVEAHGDSSTPKQLDIFVHDPRKQKT 67
QY 69 LSEKTVLTPATNMGNVTTTIPANREFKSEKGRKNFTVTVQATFTQTV-VKRVLVLSQS 127
Db 68 LFTQRYDMNPAGGMLVTPTEIPA-KEVSTDSRONQYVVVQVT-GPQVRLEKWLSSYSQS 125
QY 128 GYLFIOQDTKIYTPGSGTVLYRIPTVNNKLLPFGRTVMNENPEGIPVKQDSLSQNLG 187
Db 126 SFLFIQDTQGIYTPGSPVLYRVFSDHNTSKNKTIVIVEFQTPGEGILVSSNSVD----- 179
QY 188 VLPFSW--DIPELVNNQWKIRAYTENSQQVSTFEFEVEYVLPSPFEVTEPKFYI 245
Db 180 -LNFPPYNNLPDLVSLGTWRIIVAKYHSPEN-YTAYFDPVKYVLPSPFEVRLOPEKFFYI 237
QY 246 YNEKGLEVTITAFRLYKKEVEGTAFVIFGIQDEQRISLPESLKRIPEDSGSEVLSRK 305
Db 238 DGNENFHVSIYARYLYGEEVEGVAFVFGVKIDDAKKSIPDLSLIRIPIIDGGRATLRD 297
QY 306 VLLDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQAERSGPIVITSPQIHFHTKPKYK 365
Db 298 TFRS--RFFNLNSELVHTLYASVTYNTESGSDMVYTESGSIHIVASPYQIHFHTKPKYK 355
QY 366 PGMPFDLMVFTNPDGSPAYRVPVAVQGEDTVQSITQDGVAKLSINTHPSKPLSITVR 425
Db 356 PGMPVELTVVYVNPDPGSPAHPVYVSEAFHS-GITLSDGTAKLINLPLNAQSLPIIVR 414
QY 426 TKQELSEAEQATRTMQALPYSTVGNSSNNYHLVRLTELPGETLNINFLMRDRAHEA 485
Db 415 TNHGDLPERQATKSMATAYQTGGSGNYLHVATSTIEKPGDNLNPNVKNANSLK 474
QY 486 KIRYTYLIMNKGRLKAGRVQREPGQDLVPLSITTDFTPSFLVAVYVTLIGASGORE 545
Db 475 QIKFYIILINKGIFKVGQPRDQGNLVNHLHTPDLIPSPFRVAYQV----GNNE 530
QY 546 VVADSVWVDVKSCVSGSLVVKSGQSDROPVPGQONTLKISGDHGRVVLVAVDKGVFL 605
Db 531 IVADSVWVDVKOTCMGTLVVK--GDNLIQMPGAAMKIKLEGDPGARVGLVAVDKAVTVL 587
QY 606 NKKNKLTQSKINDVWEKADIGCTPGSGDYAGVFGSDAGLTFTSSSGQQTAKRAELQCPQ 665
Db 588 NDKYKISQAKINDTIEKSDFGCTAGSGQNNLGVFEDAGLALTSTNLNTKQESAACKCPQ 647
QY 666 AARRRR-SVOLTEKRVKVGKY-PKELRKCCEGDMRENPMRPFSCORRTFTSLGBACKKV 723
Db 648 ANRRRSSVLLDLSNASKAAEFQDQDLRKCCEGDMRENPMRPFSCORRTFTSLGBACKKV 707
QY 724 FLDCCNVITELRQARASHLGLARSLNDEDIABENIVSRSEFPESMLWNVDEL-KEPP 782
Db 708 FLECCRYIKGVRENQRESELEFLARDNEDGFIADSDIISRSDFPKSNLWLTKDLTEPN 767
QY 783 KNGISTKLWNIFLKDSITTWELAVMSDKIGICVADFEVTVMQDFIDRLPYSVVRN 842
Db 768 SQGISSTKMSFYLDRSITTWVLAVSFTPTKICVAEPYERVMKVFYFFIDQMPYSVVRN 827
QY 843 EQVEIRAVLYNRQNELKVRVELLHNPACSLATTKERHQOTITIPPKSSLSVPYIVP 902
Db 828 EQVEIRAILHNY-VNEDIYVRVELLYNPAFCASATKQGRYQQQFIKALSSRAVPFVIVP 886
QY 903 LKTGLQEVVEKAAVYHHFISDGVKSLKVYPGEGIRMNKTAVRTLDPERLEGVQKEDI 962
Db 887 LEQGLHDEIVEIKASVQBALWSDGVRKLVKVPVEGVQ-KSIVTIVKLDPRAKGVGGTQLEVI 945
QY 963 PPADLSDQVDPDTSETRILLQGTTPVAQMPEDAVDAERLKLHIVTPSGGGEQNMIGMTPTV 1022

```

Db 946 KAKLDDRPVDETEIKIIIOQBPVQAIIENSIDGSKLNHIIITPSCGGEONWIRMAAPV 1005  
Qy 1023 IAVHYLDETEQWEKFKLEKROGALELIKKGITQQLAFQPSAFAAFVKRAPSFWLTAYV 1082  
Db 1006 IATYVLDTEQWETLGINRTEAVNQIVTGYAQVVKKADSHSYAAFTNRASSSMLTAYV 1065  
Qy 1083 VKYFSLANVLLIA-IDSOVLGAVKVLILEKQKPDGVFQEDAPVHQMIGSLRNNKDM 1141  
Db 1066 VKYFAAKKAVAGISHEIICGGVRWLLNRQOPDGAFFENAPVSGTQGGIGQ-GABEEV 1124  
Qy 1142 ALTAFLVLSIQEAKDICEBVNSLPGSITKAGDFLEANYMLORSYTVAIAGYALAQWGR 1201  
Db 1125 YLTAFLVALLESKTIENDVNSLSSIKKATNLYLKKYKIQRPYTTALTAYALAAADQ 1184  
Qy 1202 LKQPLINKELTTAKDNKRNEDGKQLYNVEATSYALLALLOLKDFPVPVVRWLNQRY 1261  
Db 1185 LNDRVLMMAASTGRD--HWEENYNAHTNIEGTSYALLALLKWKPKDQGTGPIVRWLTQNF 1242  
Qy 1262 YGGYGSTOATFVFOALAQYKQADPDHQLNLDVSLQLPSSSKITHRIHWSASLRS 1321  
Db 1243 YGTYGOTQATVVAFOALAEYBIQMPHDKDNLDTIELPDREVPVRYRYNYENALLART 1302  
Qy 1322 EFTKENEGTVAEGKQGTLSVWYHAKADQOL-TCNKFDLKVITKIPAPETEKRPQDA 1380  
Db 1303 VETKLNQDITVTASGGKATMILTIFYNAQLOEKANVCNKFHLNVSE---NIHLNANGA 1359  
Qy 1381 KNTMLEICTRYRGDODATMSILDISMMTGFAPDITDLDKOLANGVDRIYSKYELDKAFSD 1440  
Db 1360 KGALMLKICTRYLGEVDSWTIIDI:SMLTGFLPDABEDLTRLKSGVDRIYSRYEDNNMAQ 1419  
Qy 1441 RNTLLIYLDKVSSEDCLAFKQVQYFENVLLTOPGAVKVAAYNLESCRTFVHPEKEDG 1500  
Db 1420 KVAVIYLNKVSSEDECHLKHFEVGFQPGSVKYSYNDLDEKCTFHPDGTG 1479  
Qy 1501 KLNLKRCDELRCBAENCFIQSDSKVTLEELERKACEPGVDYVYKTRLVKVLNSDFDE 1560  
Db 1480 LLNKICIGNVCRAGETCSLHQBRIIDVPLQIEKACETNVYVYKTRLVKVLNSDFDE 1539  
Qy 1561 YINAIQITKSGDEVOVQQRFTFSPIKREALKLEBKHYLMWGLSSDFWGEKPNLSY 1620  
Db 1540 YVMDVLEVIKQGTENPRATHQYISQRKQCALNKLKNDYVLINGSRDLPLTKDKISY 1599  
Qy 1621 IIGKDTWHERWPEDESCQENKQCODLGAFTESMVYFGCP 1662  
Db 1600 IITKNTWIERWPEDESCQENKQCODLCPDPAQPSYLTTEFGCP 1641

## RESULT 10

US-09-925-442-35  
; Sequence 35, Application US/0925442  
; Patent No. US20020103346A1  
; GENERAL INFORMATION:  
; APPLICANT: VOGEL, CARL-WILHELM  
; BREDEHORST, REINHORST  
; KOCK, MICHAEL  
; FRITZINGER, DAVID  
; TITLE OF INVENTION: RECOMBINANT PROCVF  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUBAUER,  
; P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/925,442

; FILING DATE: 10-Aug-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/017,947  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORVAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 1126-0107-0X  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1648 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 35:  
US-09-925-442-35

## Query Match

Best Local Similarity 48.7%; Score 4194; DB 9; Length 1648;  
Matches 821; Conservative 324; Mismatches 481; Indels 36; Gaps 20;

Qy 9 LLLALLLHPLALGSPMYSIITNILESEETMVLAEHDAQGDVPTVTHDFPKKLV 68  
Db 8 LVAALLIGFPGSSHGALYTLITPAVLTDTDEQILVEAHGDSPTKQLDIFVDFPRKQKT 67  
Qy 69 LSSEKTVLTATNMGNVFTTIPANREFKSEKRNKFTVQATFGTV-VEKVLVLSQS 127  
Db 68 LFQTRVDNPNAGGMLVPTTEIPA-KEVSDSQNQYVVQVVT-GPQVLEKVVLLSYQS 125  
Qy 128 GYLFIQDQKTIYTPGSTVLYRIFTVNHKLPLVGTVMNIENPEGIPVKQDLSNQNLG 187  
Db 126 SFLFTQDKGTYTPGSPVLYRVFSMDHTSKMKNKTVIBEPQTPEGLVSSNSVD----- 179  
Qy 198 VLPLSM--DIPENVMQWKIRAYENSPOQVSTEFVEKVEYVLPSEFVIVPTKFFYI 245  
Db 180 -LNFPPVNLDELVLSTGTWRAIVAKYHSPEN-VTAIFDVRKYVLPSEFVIVPTKFFYI 237  
Qy 246 YNEKGLVITITARELYGKKVEGTAFVIFIGIQDGEQRIISLPESLKRIPIEDGSGEVVLSK 305  
Db 238 DGENNFHVSITARYLYGEEVGAFLVFGVKIDDAKKSIPDSLTRIPIDGSGKATLKD 297  
Qy 306 VLLDGVQNPRAEDLVGKSLYSATVILHSGSDMVQAESGIPVITSPYOHFTKPKYFK 365  
Db 298 TFRS--RFPENLVELVHTLYASTVMTESGDMVVTESQGIHIVASPYQHFTKPKYFK 355  
Qy 366 PQMPFDLMVFTNPDGSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHPSQKPLSITVR 425  
Db 356 PGMPYELTVVTPNPDGSPAAHVPPVSEAFHSM-GTLLSDGTAKLILNPLNAQSLPITVR 414  
Qy 426 TKKQELSEAEQATETMOALPVSTVGNSSNYLHLVSLTELPCGTNNVFLMRDRAHEA 485  
Db 415 THGDLPRERQATKMTAIAVQTGGSGNVLHVATISTEIKRGDNLFPVNFVKNANSUK 474  
Qy 486 KIRYVYILMNKRLKAGQVREFGQDLVVLPLSITTDPIPSRFLVAYVYLLIGASQRE 545  
Db 475 QIKYFTYILNLNKGIFKVGQRPRDQNLVTMNLHITPDLPISFRFRVAYVQV---GNNE 530  
Qy 546 VVADSVVVDVKDSCVGLSVKSGSEDRQVPVQOQMTLKEGDHGAARVAVDKGVFLV 605  
Db 531 IVADSVVVDVDTCTNGTLVVK--GDNLIQMPGAAMKIKLEGDFGARVGLVAVDKAVYL 587  
Qy 606 NKENKLTOSKIWDVVEKADIGCTPGSGKDVAGVFSBAGLTFTTSSSQOQTAQRAELQCP 665  
Db 588 NDKYISQAKIWDITIEKSDFGCTAGSQNNLGVFEAGLALTSTNLTKQSAKCPQ 647  
Qy 666 AARRR--SVQLTREKMDKVGY-PKELRKCCEDGMENPMRFSQCRRTFISLGEACKV 723  
Db 648 ANRRRRSSVLLDSNASKAAEFQDQDKKCEDVMHNPNGYTCERAKYIQEGDACKAA 707

```

QY 724 FLDCNVITELRRCHASHGLGLARSNLDDEIIAENIUSRSFPPSSMLNVEDI-KSPF 782
DQ 708 FLECRITKGRDNGRESELSFLARDNEDGFTADSDIISRSDFPKWMLWTKOUTEBFN 767
QY 783 KNGISTKMLIFLKDSITTWELLAVMSKKGICVADPPFEVTVQDFFIDRLPYSVVRN 842
DQ 768 SQGISSTMSFYRLDSITTWVLAWSFTPKGICVAEPYEURVMKVFIDLQMPYSVVK 827
QY 843 EQVEIRAVLYNROKQELKVRVLLHNPAPCSLATTKRHHQOITIPPKSSLSVPVIVP 902
DQ 828 EQVEIRAILHNY-VNEIYVRVLLHNPAPCSLATTKRHHQOITIPPKSSLSVPVIVP 886
QY 903 LKTSQVEVEKAAVYHFIISGVKRSKLVVPEGIRMKNTVAVRTLDPERLREGVCKREDI 962
DQ 887 LEQGLHDEIKASVQEAALSDGVKRLKLVVPEGVQ-KSIVTIKLDPRAKGVGGTQLEVI 945
QY 963 PPADLSQVPTTEETRIILQGTFFVACQMTEDAVDERLHLIVTPSGCGEQNMIGWTPV 1022
DQ 946 KARKLDRVPDTEIETKIIQGPVPAQIIENSIDGSKLNLHIIITPSGCGEQNMIRMAPV 1005
QY 1023 IAVHLDTEBOWEKEGKQGLALEIKKGYTOCLAPROPSSAFAPVKRAPSTWLTAYV 1082
DQ 1006 IATYLDTEBOWEKEGKQGLALEIKKGYTOCLAPROPSSAFAPVKRAPSTWLTAYV 1065
QY 1083 KVPFLAVNLIIA-IDSOVLGAGVAKWILKQKPDGVFOEDAPVIHCEMIGGLRNNEKDM 1141
DQ 1066 KVPFLAVNLIIA-IDSOVLGAGVAKWILKQKPDGVFOEDAPVIHCEMIGGLRNNEKDM 1124
QY 1142 ALTPAVLISQBAKIDCEEOVNSLPGSITKAGDFLEANYMNGRSYTVIAIYALQAMGR 1201
DQ 1125 YLTPAVLISQBAKIDCEEOVNSLPGSITKAGDFLEANYMNGRSYTVIAIYALQAMGR 1184
QY 1202 LKGPLLNKFLITAKDKWNRDPPGKQLVNVBATSVALLLQLKDPDEVPVVRVWNEORY 1261
DQ 1185 LNDRLVLMASSTGRD--HNEYNAHFNIEGTSYALLALAKKFFQOTGPIVRWTDQNF 1242
QY 1262 YGGYSGTQATFWFQALAOYQKADPDHQLNLDVSLQPSRSSKITHRIHWSASLRS 1321
DQ 1243 YGETYGTQATFWFQALAOYQKADPDHQLNLDVSLQPSRSSKITHRIHWSASLRS 1302
QY 1322 EETKENGFTVTASGKGGLTVSVVMYHAKADOL-TCKNFDLKVTKKAPETERRPQDA 1380
DQ 1303 VETKLNQDIIVTASGKGGLTVSVVMYHAKADOL-TCKNFDLKVTKKAPETERRPQDA 1359
QY 1381 KNTWILEICTRYRQDQATWSIIDISMTGFAPTDQLKQANGVDRIYSKYELDKAFSD 1440
DQ 1360 KGALMLKICRYLGEVDSTWIIIDISMTGFAPTDQLKQANGVDRIYSKYELDKAFSD 1419
QY 1441 RNTLIIYLDKVSHEDEDLAFKHOYFNVELIQGAVKVVAYVYNNLESCETRYHPEKEDG 1500
DQ 1420 KVAIIYLNKVSHEDEDLAFKHOYFNVELIQGAVKVVAYVYNNLESCETRYHPEKEDG 1479
QY 1501 KLNKLCRDELRCRABENCFIQKSDDKVTLERLDKACBEGVDYVYKTRLVKQLSNDPDE 1560
DQ 1480 LNKICIGNVCRCAGETCSSLNHQRIDVPLQIEKACETNVDVYVYKTRLVKQLSNDPDE 1539
QY 1561 YIMAIEOTIKSGSDEVQVQORTFISPKICREALKLEKKHLMWGLSDFWGEKPNLSY 1620
DQ 1540 YVMDVLEVIKQGTDENPRAKHOYISQKCOEALNKVNDVYLINGSRSDLLPTKDKISY 1599
QY 1621 IIGKDTVWEHWPBDEECQDENKQODLGAFTESMVVFGCP 1662
DQ 1600 IITKNTWIERWPHDECEBEFQKLCDDFAQFSYTLTEFGCP 1641

```

## RESULT 11

```

US-09-925-442-34
; Sequence 34, Application US/09925442
; Patent No. US20020103346A1
; GENERAL INFORMATION:
; APPLICANT: VOGEL, CARL-WILHELM
; BREDEHORST, REINHORST

```

```

; KOCK, MICHAEL
; FRITZINGER, DAVID
; TITLE OF INVENTION: RECOMBINANT PROCVF
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/925,442
; FILING DATE: 10-AUG-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/017,947
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1126-0107-OX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1333 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-09-925-442-34

```

Query Match 41.1%; Score 3536; DB 9; Length 1333;

Best Local Similarity 51.2%; Pred. No. 9.2e-272;

Matches 683; Conservative 254; Mismatches 372; Indels 24; Gaps 15;

```

QY 334 SGSDMVQARSIGPIVTSYQ:HFTKTPKYPKPGMFDLMVFTNPDGSPAYRVPV--A 390
DQ 3 SGSDMVMTQSGIHIIVTSYQIYFTKTPKYPKPGMPEYELTVVYTKPDGSPAHPVWSEA 62
QY 391 VQGEDTVQSITQGDGVAKLSINTHPSQKPLSTTVTKQELSEAEQATRTMOALPYSTVG 450
DQ 63 IHSEGT-----TISDGTAKLFTLNTPNQASLPITVTNHDLPREQAIKSMATAYQTQ 118
QY 451 NSNNYLHSLVRLTELRPGSETLNVNFLLRMDRAHEAKIRYYTYLIMNKGRLLKAGQVREP 510
DQ 119 GSGNYLHVAITSTEIKPGDNLVFNFNVRGNANSLNQIKYFTYLLINKGKIFKVRGRHGRD 178
QY 511 GDLVLPISITTDTPSPFRLVAYTYTLIGASQREV-VADSVWVDVKDSVGSILVVKSGQ 569
DQ 179 G-NLVTWNLHIPTDLPSPFRVAYYQV----GNNEIEVADSVWVDVKDTCGTLVVKGAT 233
QY 570 SED-RQFVPGQQTMLKIEGDHGAVVLVAVDKGVFLNKKKMLTQSKLWVVEKADIGCT 628
DQ 234 SRDNRIQMFGAAMKIKLEGDPGAWIGLVAVDKABEVLNDKDKYISQAKIWDTIKSDFGCT 293
QY 629 PGSGKDYAGVFS DAGLTFTSSSGQQTQAPRAELQCQPPAARRR-SVQLTERKMDKGVKY- 686
DQ 294 AGSGQNNLGVFEDAGLALTSTNLTNKQSAKCCQPPANRRRRSVLLDSNASKAAQFQ 353
QY 687 PKELRCCBEGMRENPMRFSQQRTRFISLGRACKKVFLLDCQNYITELRRQHAASHLGL 746
DQ 354 DQDLRCCBEGMRENPMRFSQQRTRFISLGRACKKVFLLDCQNYITELRRQHAASHLGL 746

```





QY 1594 FISPIKREALKLEEKHYLMWGLSSDFWGEKPNLSYIIIGKDTWVEHWPEDCCDEBNQ 1643  
Db 626 FISPIKREALKLEEKHYLMWGLSSDFWGEKPNLSYIIIGKDTWVEHWPEDCCDEBNQ 685  
QY 1644 KOCQDLGAFSTSMVVFQCPN 1663  
Db 686 KOCQDLGAFSTSMVVFQCPN 705

## RESULT 13

US-09-842-758-6  
; Sequence 6, Application US/09842758  
; Publication No. US20030083244A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A. M.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Spytke, Kimberly A.  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Rastelli, Luca  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Taupier, Raymond J.  
; APPLICANT: Grosse, William M.  
; APPLICANT: Edward, Szekeres S.  
; APPLICANT: Alsobrook II, John P.  
; TITLE OF INVENTION: NO. US20030083244A1el Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-783  
; CURRENT APPLICATION NUMBER: US/09/842,758  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/200,158  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: 60/200,613  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,780  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/201,006  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,007  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,236  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,238  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/201,186  
; PRIOR FILING DATE: 2000-05-02  
; PRIOR APPLICATION NUMBER: 60/201,474  
; PRIOR FILING DATE: 2000-05-03  
; PRIOR APPLICATION NUMBER: 60/201,508  
; PRIOR FILING DATE: 2000-05-03  
; PRIOR APPLICATION NUMBER: 60/220,591  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: 60/232,678  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: 60/263,217  
; PRIOR FILING DATE: 2001-01-22  
; PRIOR APPLICATION NUMBER: 60/265,160  
; PRIOR FILING DATE: 2001-01-30  
; NUMBER OF SEQ ID NOS: 113  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1612  
; TYPE: PRT  
; ORGANISM: Homo sapiens

; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1226)  
; OTHER INFORMATION: Wherein Xaa is any amino acid.  
US-09-842-758-6  
Query Match 33.2%; Score 2857.5; DB 10; Length 1612;  
Best Local Similarity 36.8%; Pred.No. 1.6e-217;  
Matches 644; Conservative 314; Mismatches 562; Indels 231; Gaps 39;  
QY 1 MGPTSGPSLLLLLLTHPLALGSPMYSIITENILRESEETWLEAHDAQGD-----VFV 55  
Db 1 MSPSLPFSALISLCLSHLPWLPSPRYILVTPRLVAVGSPESIHIOAHSDSQPLRRLTKV 60  
QY 56 TVTVHDFPGKGLVLSSEKTVLTPATHMGNVTFTIPANREFEKSEKGRNKFVTVQATF--- 112  
Db 61 NLTVWDFPMKRTVILARSQILSPGNFMDDQAPVTPESLMYLDLPQGQ-QYIILRATWAPT 119  
QY 113 -GTQVVEKVLVLSQSGYLFIQTDKIITYPGSTVLRYIFTVNHKLPLVGRTWVNIENPE 171  
Db 120 SGSSFMEKMWLVALHAGYIFIQTEKTIYTPSLVHYRVFTVNHKMDPVTRITFLDKNPD 179  
QY 172 GIPVKQDSLSSQNLGLVPLSWDIBP---LVNMGMOKIRAYVENSPOOVFSTEFVEKVV 228  
Db 180 GSPASRVLVHSQQQPGVGLASWG-PEKSLCLSGTWTIERASYQSTPKKFAAFDVKVY 238  
QY 229 LPSFEVIVPTEKPYIYNEKGLVITITARFLYKKVEGTAPVIFIGIQDQORISLPESL 288  
Db 239 LPSFEVQLVFNKTFYFLKDE-ALGVDIQARYIFNKPVDGHALVIFGVKLDSCRPIQSSL 297  
QY 289 KRIPIEDGSGEVVLSKVLDDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQAESGPI 348  
Db 298 QRVEVTE-----TGGEMVQAESGVKI 319  
QY 349 VTSPIQIHFTKTKPKPKQMPDL-----M 373  
Db 320 IQSPYNIKTRTPQYFKPGMPHFRRVQSSPIQIIFQSHLSHOATAGSFLLPQIPBP 379  
QY 374 VFVTNPDGSPAYRVPVAVOGEDTVQSLTQGDGVAKLSINTHPSQKPLSITVTRTKQELSE 433  
Db 380 VFIENPDGSPASRVLVHSQDQ---KYVTSAGELATLINTDANLKGFLIEVKT-EESLQP 435  
QY 434 AEOATRTMQALPYSTVGNNSNYLHLSV--LRTELPGETLVNVLRLVRDRAHEAKIRYTT 491  
Db 436 EQASAKMTAWPYLTQDGSNFLHIEVKTGLTEVGSSIQLSLN-TRHQDPKTKDKITHT 494  
QY 492 YLIAMNKGRLKAGQVRPEPQDLVPLSLITTDFTPSRLVAYVTL-TGASQQRVWADS 550  
Db 495 ILVREGKARQGRQVAGV-----VPSRILAFYLLPRASQODELVADS 540  
QY 551 VWVDVKDSCVGSVLWKSQSEDR---QVPVGGQMTLKEGDHGRVVLVAVDKGVFLNKK 607  
Db 541 IWIDVNDRCIG--LVKGLKNDRFFQSLPNSQVELKVTGDABATVGLVAVDKAVYVLS 597  
QY 608 KXKLTQSKTDVVEKADIGCTFGSGKDYAGVFSAGLFTFTSSGQOQTAQRAELOCP--- 663  
Db 598 KHLTKQKVMNVVEHDICTGSGKDRPAVFKDAGLDLKISTGMDSHQSHSCQEAEV 657  
QY 664 -----QPAARRRSVQTEKMDKVGYKPKEL-RKCCEDGMRENPMRFQCRTRPISIGE 718  
Db 658 GESLEPGROKRLRSAAETAPLH-SVNVNFKTELEKQCEAGLRESVPLGSCERTWVRHGP 716  
QY 719 ACKKVFLLDCNXYITELRRQHARASHLGLARSLNLDDEIIAEENIVSRSEFPESLWNVDEL 778  
Db 717 ACVAFLDCCSHL-----LPPADEEDFDDLFLDDMPVRTLFPESWLW----- 759  
QY 779 KEPPKNGISTKLMNIFLKDSITTWELAVSM-SDKKGI CVADPPFVTVMQOFFIDRLRPY 837  
Db 760 -----NSISHYPISTVKVPDSITTFQVWVSLKAGQGLCVSDPPELTVMKFEFFVLDKLP 814  
QY 838 SVVRNEQVETRAVLVYNRQNELKYRVVLLHNPAFCSLATTKRRHQQTITIPKSSLSVP 897  
Db 815 SVIRNEQVQIQAMLVNFDRDQ-AKRVVEPFPKTEILCSASKPGAPSHQVTVVPFSSKLVH 873



Db 239 LBSFEVQVLPNKTFFYLKDE--ALGVDIOARYIFNKPVDGHALVIFGVKLDSCRIPIQSSL 297  
QY 289 KRIPIEDSGEVLRSKRYLLDGVONPRAEDLVGKSLVSVATVILHSGSDMWQAERSGIP 348  
Db 298 ORVEYTE-----TGEMVQAETSGVKI 319  
QY 349 VTSPOVQIHFTTKYKFGMPFDL-----M 373  
Db 320 IQSPNIAFTTTPQVFRGMPFHRVRVVOSSPIQIIFQSHLHQATAGSFLLPIQPPQ 379  
QY 374 VFTVNPDSGARVPVAVQSGEDTQVOSIQDGVAKLSINTHPQCKLSITVTRKQELSE 433  
Db 380 VFSNPDGSPASRVLVHSQDQ---KVYTSAGLATLTINTDANLKLPIEVKT--BESLOP 435  
QY 434 BQCATRTWALPYSVTGNSNNYHLSV--LETFLPCGTINLVNFFLRMDRAHEAKIRYX 491  
Db 436 BQASAKWTANPYLTQDQSGNLFHEVTKLTGTEVGSIIQSLN--TRHQPKYKDKHTFT 494  
QY 492 YLIMNKGELLKAGQVREPQDQVVLPLSTITPISFRLVAYTL--IGASGOREVADS 550  
Db 495 ILVWREGKARGLGRQVAGV-----VPSFRILAFYLLPRGASQDPELVADS 540  
QY 551 VWVDVKDCVGLSVVKSQSEDR---OPVPCQQWTLKIEGDHGARVVLVAVDKGVFLANK 607  
Db 541 IWDVNDRCIG--LKVGLKNDPRFQSLPNSQVELKVTGDAEATVGLVAVDKAVYVLS 597  
QY 608 KXKLTQSKIWDVADKADICTPGSGKDYAGVFSAGLTFTSSGQQTAAQRAELQCP---- 663  
Db 598 KXKLTQKVVNVNBEHDIGCTGGSGKDRFAVFKDAGLDLKIETCMBDSGHQSSCOEAEV 657  
QY 664 ---OPAAARRRSVQLTEKMDKVKYKPKEL--KKCCEDGMRENPMFSCQRRTRFISLGE 718  
Db 658 GESLEPGQRILRSIEIAPLH--SSVYKPKTELEQKCCAGLRESPVGLSCSERTWVTHP 716  
QY 719 ACKVFLDCCNYITELRQARASHGLARSNLDPIIAEENIVSRSEPPESLMNVDEL 778  
Db 717 ACVAFLDCCSGL-----LPPADEEEDFDLDDMPVTRTLFPESLW----- 759  
QY 779 KEPPKNGSTKLMNIFLKDSITWELLAVSN--SDKGGICVADPFVTVWCDRFDLRLPY 837  
Db 760 -----NSISHYPISVKVPDSITWQVVSLLKAGQGLCVSDPFEITWMSFEVDLKLPS 814  
QY 838 SVRNEQVEIRAVLYNRQNELKRVVELLHNPFCSLATKRHOCTITIPKSSLSVP 897  
Db 815 SVIRNEQVQIOAMLYNFRDQ--AKRVVEFFHETKLTCSAKPGAPSHQVWVVPPTSSKIVH 873  
QY 898 YVIVPLKTLGOEVEKAAVYHFTSDGVRKSLKVPVREGIRMNKTAVRTLDPERLREG- 956  
Db 874 FVLPLETKVDVEKANGYG--VQDHVKKTLV-----EGCYISQ 913  
QY 957 VOKEDI PADLSQDVPDTESSTRILLQGTVPVQKTEDAVDAERLKHILVTPSCGQNM 1016  
Db 914 TQTKLVPRQEFILNMVPTDEAEVFSVQGDILGETIVGSLTPSPBIOQLLRVPTGCPQTLS 973  
QY 1017 GMTPTVIAVHYLDETEQWEKEGLEKROGALELIKGYTQOLAFRQPSAPAFVKEAPST 1076  
Db 974 SLTPVILSLRVLDITGQKGVGVRHQVKNL--GYTQMLTRSSDGYHTSKNGPST 1031  
QY 1077 WUTAVVVKFSLAVNLI---AIDSVLCGAVKMLILEKQKPDGVQEDAPVHQEMIGSL 1133  
Db 1032 WUTSVFVFALAYSMMTQVLSLSLDCDMMANMIIIDQAEADGHFLEKGPVVMVTS- 1087  
QY 1134 RNNKEDVALTAVLISLOEAKDICEBOVN--SLPGSITKAGDPLEANYMNLORSYVATA 1192  
Db 1088 ---SEEDVSLTALVLIANEKELCRQKVGPNLMASIEKAGGLELRLRLQESYAVATA 1144  
QY 1193 GVALAQMLKGLPLINKLFTTAKOKNRW---EDPGKOLYNVEATSVALLALQLKDPDFV 1249  
Db 1145 SYALA-----DKTHWPVDEQNLGSLTYTIEATAYLQKLELGRYNET 1186  
QY 1250 PVVRWMLNEORYGGVGSST-----QATFVWFOALACYOKDAPDHQJELN 1294  
Db 1187 HAIKXVLEKELGGLGGFSTPGSSRLSHSPQRWPGSLXALALTRFRAVFPKGIQDL 1246

QY 1295 DVSLQPLSRSSKITHRIHWESALLRSEETKENEGFTVTAEGKQOGLSVVWTHYHAKAD 1354  
Db 1247 HVQIRAPKATLVNWNVIDHNSAYQOSAKFLAQDLDLEIKASNGRGTISILTYHKSPE 1306  
QY 1355 -OLTCNKFIDLKVTIKAPETEKEPQDAKNTMILEICTRYRGQDQATMSLTDISMMTGPAP 1413  
Db 1307 REDCNLYHLNATLHSALEENK--GGETFLRMETRFQNNGEATMTIMEVSLTGYPP 1363  
QY 1414 DTDDLKOLANGVDRYISKYELDKAFSDRNTLIIYDKVSHSDEDDCLAFKVOHYFVNELIQ 1473  
Db 1364 NQDDLKOLTSDVERVAFQYKTKTSTSD--STVYLYLEKLKSHKNTVLGFRVHRMLQAEFLQ 1422  
QY 1474 PGAVKVVAYNLEESCTRFVHPEKEDGKLNKCRDLCPCAEENC--FTQKSDDKVTLBER 1532  
Db 1423 AALVTIYDYEPRRSGSTFYNLPTQESSLRKCHKDICCASCQCFPSLOXPSGQLRQBEL 1482  
QY 1533 LDKACEPGVDYVYKTRL--VKVQLSNDFFEYIMAEQITKSGSDEVQVQOQRTFISPKC 1590  
Db 1483 QTTACEAGVDVYKTKLESVEVSASNPYVYNTQLEDIIKSGTDPKAPLAKKKFVSHATC 1542  
QY 1591 REALKLEEKHYLMWGLSSDFWCEKENLSYVIIGKDTWVHWPBEECCQDEENKQCOQDLG 1650  
Db 1543 HDSLGLGEQSYLIIMQTSDLWRIKSDYSVVLGKEFTLLWADGASKKELRDLQEE-- 1600  
QY 1651 AFTESWVVFEC 1661  
Db 1601 -FLEYNRTHGC 1610

RESULT 15  
US-09-981-151A-80  
; Sequence 80, Application US/09981151A  
; Publication No. US20030212256A1  
; GENERAL INFORMATION:  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: MacDougall, John R  
; APPLICANT: Malyankar, Muriel M  
; APPLICANT: Smithson, Glenada  
; APPLICANT: Mullet, Isabelle  
; APPLICANT: Peyman, John A  
; APPLICANT: Stone, David J  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Taupier Jr, Raymond J  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spvtek, Kimberly A  
; APPLICANT: Gangolli, Eata A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Gorman, Linda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-168  
; CURRENT APPLICATION NUMBER: US/09/981,151A  
; CURRENT FILING DATE: 2001-10-16  
; PRIOR APPLICATION NUMBER: 60/241,040  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/241,058  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/241,063  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/241,243  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/242,152  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/242,482  
; PRIOR FILING DATE: 2000-10-23

Db 708 KVSHSETECVGFKHQDFEVLGLOPASVKYDIYEDQCTAFY 751

Search completed: December 22, 2004, 00:49:58  
Job time : 199 secs

```

; PRIOR APPLICATION NUMBER: 60/242,611
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,612
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/242,880
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 60/242,881
; PRIOR FILING DATE: 2000-10-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 751
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: Alpha-2-macroglobulin family Consensus Sequence
US-09-981-151A-80

Query Match      21.6%; Score 1855.5; DB 10; Length 751;
Best Local Similarity 52.4%; Pred. No. 3.2e-138;
Matches 400; Conservative 105; Mismatches 220; Indels 39; Gaps 10;

Qy 756 IAEENIVSRSPESLWVDELKPPKNGISTKLWNFLKDSITTWELLAVSNMDDKGI 815
Db 1 IDEDDITIRSYFFPSLWVEVEVDRSP-----VLTVNITLPSDITTWELLAVSNMDDKGI 55

Qy 816 CVADPEVITVQDFIDLPLYSVVRNEQVEIRAVLYNQONQELKVRVELLHNP---A 871
Db 56 CVADPEVITVQDFIDLPLYSVVRNEQVEIRAVLYNQONQELKVRVELLHNP---A 871

Qy 872 FCSLATKRRHQTIIPKSSLSVYVIVPLKTGLQVEVKAAYVHHFISDGVKRSIKV 931
Db 116 FCSLATQRTSSQSVKPSLSSVSVVYVPLASGLSLVEVVASVPEFFVKDAVVKLKV 175

Qy 932 VPEGIRMNKTVAVRTLDPRLREGVQKEDIPPAQLSDQVDPDTESETRILLQGTPTVAQMT 991
Db 176 EPGARKEETVSLLLPPEHLG-GGLEVSEVPALKLPDDVPDTEAEAVISVQGDPTVAQAI 234

Qy 992 EDAVDAERLKHIVTSGGEGQNMGTPTVIAYHVLDETEQWEKFGLEKQGALELIK 1051
Db 235 QNTLSGEGNLLRLPSGGEGQNMGTPTVIAYHVLDETEQWEKFGLEKQGALELIK 294

Qy 1052 GYTQQLAFRQPSAFAFVKRAPSTWLTAYVVKVFLAVNLTAIDSVLCGAVKWLILEK 1111
Db 295 GYQRLNRYKADGSYAAFLHRASSTWLTAFVLKVFSQARNYVVIDEBHICGAVKWLILNQ 354

Qy 1112 QKPDGVPOEDAPVTHQEMIGGLNNEKDMALTAFVLISLOEAKDICEQVNSLPGSITK 1171
Db 355 QKDDGVFREGSPVHNEMKGGVGDAAVEVTLTAFITALLLEAKLVCSIPVAVNALSILK 414

Qy 1172 AGDFLEANYNLOQSVTVAIAGYALAQMG---RLKGPL--LNKFLTTAKDNRMEDPKG- 1225
Db 415 ASDYLLNRYANGQVVTTLTAYALAGVLHKLKEILSKLEELYKALVKGHWERPQKP 474

Qy 1226 -----QLYNVEATSYALLALLOLKDF---DFVPPVYVWLNQRYGGYGSTQ 1270
Db 475 KDAFPHYSPQQAAYAVEMTSYALLALLTLFPKVEKAPKVVKWLTEQQYGGYGSTQ 534

Qy 1271 ATFWVQALAQYQKADAFDQELNLDVSLQLPSSKSIKTHRIHESASLLRSEETKENE 1330
Db 535 DTVMALQALSKYGIATPTHKENKLSVTIQSPSGFSKSHFQILNNAFLRLRPVELPLNEGF 594

Qy 1331 TVTAEGKGGTGLVTVWYHAKAKDQLTCKNFKLVTIKPAPETEKRPQDAXNTMLBCT 1390
Db 595 TVTAKVTGQGLTILVTYRYKLDKXNTFCFDLK--IETVPTICVEPKAKNSDYLSICT 652

Qy 1391 RYRGDQ-DATWSILDISMNTGPAFPTDDLLKQLANGVDRIYSKYELDKAFSDRNTLIYLD 1449
Db 653 RYAGSRSDGMALADISMLTGFIPLKPLKLENGVDRIYSKYEI-----DGNHVLAYLD 707

Qy 1450 KVSHSEDDCLAFKVHQYFVNVLLQPGAVKVYAVYNLEESCTRFY 1493

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2004, 00:29:54 ; Search time 45 Seconds  
(without alignments)  
2450.819 Million cell updates/sec

Title: US-09-875-519A-22

Perfect score: 8609  
Sequence: 1 MGPTSGPSLLLLLTHPLA.....KQCQDLGAFTEMSVVGCPN 1663

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCITUS COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfileesi.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8609	100.0	1663	2	US-08-793-126-1
2	8609	100.0	1663	3	US-09-132-271-1
3	8609	100.0	1663	3	US-09-142-334-22
4	4389.5	51.0	1651	1	US-08-447-411-2
5	4224	49.1	1642	2	US-08-662-227-2
6	4224	49.1	1642	3	US-09-017-947-2
7	4224	49.1	1642	4	US-09-925-442-2
8	4208	48.9	1642	1	US-08-447-411-45
9	4194	48.7	1648	2	US-08-662-227-35
10	4194	48.7	1648	3	US-09-017-947-35
11	4194	48.7	1648	4	US-09-925-442-35
12	3359	41.3	1333	1	US-08-447-411-76
13	3336	41.1	1333	2	US-08-662-227-34
14	3336	41.1	1333	3	US-09-017-947-34
15	3336	41.1	1333	4	US-09-925-442-34
16	3255	37.8	645	1	US-07-779-172A-3
17	1840	21.4	1676	3	US-08-487-283A-2
18	1287.5	15.0	308	4	US-09-582-761B-26
19	1287.5	15.0	330	4	US-09-582-761B-37
20	1287.5	15.0	929	4	US-09-582-761B-27
21	1057	12.3	403	1	US-08-118-674-1
22	1016	11.8	1474	4	US-09-241-606-2
23	659	7.7	134	4	US-09-513-999C-4153
24	415	4.8	77	4	US-08-708-606-1
25	407	4.7	76	1	US-08-284-022-1
26	407	4.7	77	3	US-09-246-500B-1
27	339	3.9	63	1	US-08-447-411-24

28	339	3.9	63	1	US-08-447-411-63	Sequence 63, Appl
29	339	3.9	63	2	US-08-662-227-20	Sequence 20, Appl
30	339	3.9	63	3	US-09-017-947-20	Sequence 20, Appl
31	339	3.9	63	4	US-09-925-442-20	Sequence 20, Appl
32	314.5	3.7	78	3	US-09-246-500B-5	Sequence 5, Appl
33	302	3.5	63	1	US-08-447-411-26	Sequence 26, Appl
34	301	3.5	63	1	US-08-447-411-25	Sequence 25, Appl
35	291	3.4	63	1	US-08-447-411-27	Sequence 27, Appl
36	288	3.3	77	3	US-09-246-500B-2	Sequence 2, Appl
37	288	3.3	106	2	US-08-569-186-33	Sequence 33, Appl
38	283	3.3	317	4	US-09-270-767-43623	Sequence 43623, A
39	282.5	3.3	78	3	US-09-246-500B-3	Sequence 3, Appl
40	276.5	3.2	78	3	US-09-246-500B-4	Sequence 4, Appl
41	203	2.4	339	4	US-09-270-767-45607	Sequence 45607, A
42	199	2.3	250	4	US-09-241-606-4	Sequence 4, Appl
43	187	2.2	145	4	US-09-270-767-59001	Sequence 59001, A
44	186	2.2	63	1	US-08-447-411-23	Sequence 23, Appl
45	183.5	2.1	3878	4	US-09-914-259-11	Sequence 11, Appl

## ALIGNMENTS

RESULT 1  
US-08-793-126-1  
; Sequence 1, Application US/08793126  
; Patent No. 5843297  
; GENERAL INFORMATION:  
; APPLICANT: Harrison, Richard Alexander  
; APPLICANT: Farries, Charles Timothy  
; TITLE OF INVENTION: MODIFIED HUMAN C3 PROTEINS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE AND DORR LLP  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: United States of America  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/793,126  
; FILING DATE: 07-FEB-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baker, Hollie L.  
; REGISTRATION NUMBER: 31,321  
; REFERENCE/DOCKET NUMBER: 102286.377  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 526-6000  
; TELEFAX: (617) 526-5000  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1663 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-793-126-1

Query Match 100.0%; Score 8609; DB 2; Length 1663;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGPTSGPSLLLLLTHPLALGSPMYSITPNILRLESEETWLEADAGDVPTVTH 60

Db 1 MGPTSGPSLLLLLTHPLALGSPMYSITPNILRLESEETWLEADAGDVPTVTH 60

Qy 61 DFFGKLVLSSEKTVLTPATNMGNVTFTIPANREFKSEKGRNKFVTQATFGQVVEKV 120

Db 61 DFPCKLVLSSEKTLTPATNMGVNTFTIPANREFKSEKGRNKFVTVQATFGTVVEKV 120  
QY 121 VLVSLOSGLFIQDXTIYTPGSTVLRYFTVNHKLLPVGRVWNIENPESGIVKQDSL 180  
Db 121 VLVSLOSGLFIQDXTIYTPGSTVLRYFTVNHKLLPVGRVWNIENPESGIVKQDSL 180  
QY 181 SSQNLGVLPLSWDIPELVNMGMKIRAYENSPQOQFSTEFVEKYYVLPSEFVIEPTE 240  
Db 181 SSQNLGVLPLSWDIPELVNMGMKIRAYENSPQOQFSTEFVEKYYVLPSEFVIEPTE 240  
QY 241 KFYIYNEKLEVTIITARFLYKKGVEGTAFVIFGQDGEORISLPSLESKRPIEDGSGEV 300  
Db 241 KFYIYNEKLEVTIITARFLYKKGVEGTAFVIFGQDGEORISLPSLESKRPIEDGSGEV 300  
QY 301 VLGRKVLLDGVQNPRAEDLVGSLYVSATVILHSGSDMVQAESGIPVTSPIQHTKT 360  
Db 301 VLGRKVLLDGVQNPRAEDLVGSLYVSATVILHSGSDMVQAESGIPVTSPIQHTKT 360  
QY 361 PKYFKGMPDLMVFTNPDGSPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSOKPL 420  
Db 361 PKYFKGMPDLMVFTNPDGSPAYRVPVAVQGEDTVQSLTQDGVAKLSINTHPSOKPL 420  
QY 421 SITVRTKQSLSEABOATRTMQALPYSTVGNSSNNYHLVSRTLRPGETLNVLNFLRMD 480  
Db 421 SITVRTKQSLSEABOATRTMQALPYSTVGNSSNNYHLVSRTLRPGETLNVLNFLRMD 480  
QY 481 RAHEAKIRYTYILMKNKRLKAGROVRPQDQVLVPLSITTFIPSFRLVAVYTLIGA 540  
Db 481 RAHEAKIRYTYILMKNKRLKAGROVRPQDQVLVPLSITTFIPSFRLVAVYTLIGA 540  
QY 541 SQBEVVADSVWVDVDSKCSVLVKSQSESDRQVPGQOMTLKIEGDHGRVVLVAVDK 600  
Db 541 SQBEVVADSVWVDVDSKCSVLVKSQSESDRQVPGQOMTLKIEGDHGRVVLVAVDK 600  
QY 601 GVFVLNKKNLTKOSKIDWVVEKADICTPGSGKDYAGVFSAGLTFTTSSSQOQTAQRAEL 660  
Db 601 GVFVLNKKNLTKOSKIDWVVEKADICTPGSGKDYAGVFSAGLTFTTSSSQOQTAQRAEL 660  
QY 661 QCPQAAARRRSVQLTEKMDKVGKPKELRCCEDGMRENPMRPFSCORRTFTSLGEAC 720  
Db 661 QCPQAAARRRSVQLTEKMDKVGKPKELRCCEDGMRENPMRPFSCORRTFTSLGEAC 720  
QY 721 KKVFLDCCNVITELRQHARASHGLARSNLDEDIIAENIVSRSEFPESLMWVDELKE 780  
Db 721 KKVFLDCCNVITELRQHARASHGLARSNLDEDIIAENIVSRSEFPESLMWVDELKE 780  
QY 781 PKNGISTKLMNIFLKDSTITWEILAVNSDKKICVADPFEVTVMDQFIDILRLPVSV 840  
Db 781 PKNGISTKLMNIFLKDSTITWEILAVNSDKKICVADPFEVTVMDQFIDILRLPVSV 840  
QY 841 RNEQVEITRAVLYNRQNELKVRVLLHNPACSLATTKRRHQOTITIPPKSSLSVPYVI 900  
Db 841 RNEQVEITRAVLYNRQNELKVRVLLHNPACSLATTKRRHQOTITIPPKSSLSVPYVI 900  
QY 901 VPLKTGLQVEVKAAYVHHFISDGVKSLVVPVPEGRVNNKTVAVRTLDPRELGRGVQKE 960  
Db 901 VPLKTGLQVEVKAAYVHHFISDGVKSLVVPVPEGRVNNKTVAVRTLDPRELGRGVQKE 960  
QY 961 DIPPADLSQVPTSETIRILQGTPTVAQMTEDAVDAERLKLIVTPSGCGEQNMIQWTP 1020  
Db 961 DIPPADLSQVPTSETIRILQGTPTVAQMTEDAVDAERLKLIVTPSGCGEQNMIQWTP 1020  
QY 1021 TVIAVHYLDTEQWEXFLEKQGALELIKKGYTQQLAFROPSSAFKVRAPSTWLTA 1080  
Db 1021 TVIAVHYLDTEQWEXFLEKQGALELIKKGYTQQLAFROPSSAFKVRAPSTWLTA 1080  
QY 1081 YVVKVFLAVNLIAIDSVLCGAVKWLIILEKQKPGVQEDAPVHQBIMIGLNNREKD 1140  
Db 1081 YVVKVFLAVNLIAIDSVLCGAVKWLIILEKQKPGVQEDAPVHQBIMIGLNNREKD 1140  
QY 1141 MALTAFLVLSIQBAKICEQVNSLPGSITKAGDFLEANNMQLORSYTVATAGYALQMG 1200

Db 1141 MALTAFLVLSIQBAKICEQVNSLPGSITKAGDFLEANNMQLORSYTVATAGYALQMG 1200  
QY 1201 RLKGPLLNKFLTTAKQKNRWEDFGKOLYXNVEATSVALLALLQKDFDVPVVRWLNQOR 1260  
Db 1201 RLKGPLLNKFLTTAKQKNRWEDFGKOLYXNVEATSVALLALLQKDFDVPVVRWLNQOR 1260  
QY 1261 YYGSGYSTQATFWFOALAQYOKDAPDHOELNVLVSLQPSRSSKITHRHWSASLLR 1320  
Db 1261 YYGSGYSTQATFWFOALAQYOKDAPDHOELNVLVSLQPSRSSKITHRHWSASLLR 1320  
QY 1321 SEETKENEGFTVTAEGKGQGLSVWTHYHAKAKDQLTCKNFKDLKVTIKPAPETEKRPQDA 1380  
Db 1321 SEETKENEGFTVTAEGKGQGLSVWTHYHAKAKDQLTCKNFKDLKVTIKPAPETEKRPQDA 1380  
QY 1381 KNTWILICITRYRGDQDATWSILDIISMTGFPADPTDDDLKOLANGVDRYISKYELDKAFSD 1440  
Db 1381 KNTWILICITRYRGDQDATWSILDIISMTGFPADPTDDDLKOLANGVDRYISKYELDKAFSD 1440  
QY 1441 RNTLIIIVLDKXVSHSEDDCLAFKVHQYFNVNELIQPGAVKYVYAYNLEESCTRFYHPEKEDG 1500  
Db 1441 RNTLIIIVLDKXVSHSEDDCLAFKVHQYFNVNELIQPGAVKYVYAYNLEESCTRFYHPEKEDG 1500  
QY 1501 KLNKLCDELCRCABENCFIQKSDDDKVTLLEERLDKACBPGDYVYKTRLVKQVLSNDPDE 1560  
Db 1501 KLNKLCDELCRCABENCFIQKSDDDKVTLLEERLDKACBPGDYVYKTRLVKQVLSNDPDE 1560  
QY 1561 YVMAEQTIKSGSDEVQVQGGQRTFISPIKREALKLEKKHYLMWGLSSDFWGEKPNLSY 1620  
Db 1561 YVMAEQTIKSGSDEVQVQGGQRTFISPIKREALKLEKKHYLMWGLSSDFWGEKPNLSY 1620  
QY 1621 IIGKDTWVHEWPEDEBCQDEENKQCODLGAFTESMVWFGCPN 1663  
Db 1621 IIGKDTWVHEWPEDEBCQDEENKQCODLGAFTESMVWFGCPN 1663

## RESULT 2

US-09-132-271-1  
Sequence 1, Application US/09132271  
Patent No. 6221657  
GENERAL INFORMATION:  
APPLICANT: Harrison, Richard Alexander  
APPLICANT: Faries, Charles Timothy  
TITLE OF INVENTION: MODIFIED HUMAN C3 PROTEINS  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE AND DORR LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: United States of America  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/132,271  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/793,126  
FILING DATE: 07-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Baker, Hollie L.  
REGISTRATION NUMBER: 31,321  
REFERENCE/DOCKET NUMBER: 102286.377  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 526-6000  
TELEFAX: (617) 526-5000  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1663 amino acids

; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-09-132-271-1

Query Match 100.0%; Score 8609; DB 3; Length 1663;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPTSGPGLLLTHLPALGSDWYSIIITNIRLSEBETMULEAHDAQGVPTVTVH 60  
 DB 1 MGPTSGPGLLLTHLPALGSDWYSIIITNIRLSEBETMULEAHDAQGVPTVTVH 60  
 QY 61 DFPKGLVLSSEKTVLTPATNMGNTFTIPANREPKSEKRNKFTVQATGTVQVEKV 120  
 DB 61 DFPKGLVLSSEKTVLTPATNMGNTFTIPANREPKSEKRNKFTVQATGTVQVEKV 120  
 QY 121 VLVSLSQGVLFQTDKTIYTGSTVLYRIFTVNHKLLPVGRTVMNINPEGIPVKQDSL 180  
 DB 121 VLVSLSQGVLFQTDKTIYTGSTVLYRIFTVNHKLLPVGRTVMNINPEGIPVKQDSL 180  
 QY 181 SSQNLGVLPDLSWIDPELVNNGQWKIRAYYENSPOQVFSTPEPVKXYVLPSEVIVEPTE 240  
 DB 181 SSQNLGVLPDLSWIDPELVNNGQWKIRAYYENSPOQVFSTPEPVKXYVLPSEVIVEPTE 240  
 QY 241 KFYIYNEKGLVITITARFLYKKGVEGTAFFVIGIQDGEORISLPSLXRIPIEDSGSEV 300  
 DB 241 KFYIYNEKGLVITITARFLYKKGVEGTAFFVIGIQDGEORISLPSLXRIPIEDSGSEV 300  
 QY 301 VLRSKVLIDGVNPRADLVGKSLYVSATVILHSGSDMWQAESGPIVTSYQHFHTKT 360  
 DB 301 VLRSKVLIDGVNPRADLVGKSLYVSATVILHSGSDMWQAESGPIVTSYQHFHTKT 360  
 QY 361 PKYFKPGMPFDLMVFVTPNDCSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHSQKPL 420  
 DB 361 PKYFKPGMPFDLMVFVTPNDCSPAYRVPVAVQGEDTVQSLTQGDGVAKLSINTHSQKPL 420  
 QY 421 SITVTKKQELSEAEQARTWQALPYSTVGNNSNYHLVRLTELPGETLVNHFLLRMD 480  
 DB 421 SITVTKKQELSEAEQARTWQALPYSTVGNNSNYHLVRLTELPGETLVNHFLLRMD 480  
 QY 481 RAHEAKIRYTYILMNKGRLLKAGQVREPQDLVPLSITTDFTFSPRLVAYVTLIGA 540  
 DB 481 RAHEAKIRYTYILMNKGRLLKAGQVREPQDLVPLSITTDFTFSPRLVAYVTLIGA 540  
 QY 541 SCOREVADSVVQVDCSVGLVYKSGQSDRQPVGQOMTLKIEGDHGAIVLVANDK 600  
 DB 541 SCOREVADSVVQVDCSVGLVYKSGQSDRQPVGQOMTLKIEGDHGAIVLVANDK 600  
 QY 601 GFVNLKXNKLTQSKINDVWEKADIGCTPGSGKDYGAVFSDAGLTFTSSGGQQTARABL 660  
 DB 601 GFVNLKXNKLTQSKINDVWEKADIGCTPGSGKDYGAVFSDAGLTFTSSGGQQTARABL 660  
 QY 661 QCQPAARRRSVOLTEKMDKVKYKPKELRKCCEGDMRENPNRFSQCRRTFISLGEAC 720  
 DB 661 QCQPAARRRSVOLTEKMDKVKYKPKELRKCCEGDMRENPNRFSQCRRTFISLGEAC 720  
 QY 721 KKVFLDCCNYITELRQCHARASHLGLARSLNDEIIAENIVSRSEFPESLWNVEDLKE 780  
 DB 721 KKVFLDCCNYITELRQCHARASHLGLARSLNDEIIAENIVSRSEFPESLWNVEDLKE 780  
 QY 781 PPKNGISTKLMMIFLKDSTITWEILAVMSDKKIGICVADPFVTVMQDFIDLRLPYSVV 840  
 DB 781 PPKNGISTKLMMIFLKDSTITWEILAVMSDKKIGICVADPFVTVMQDFIDLRLPYSVV 840  
 QY 841 RNEQVEIRAVLYNFRQNELKVRVELLHNPAFCSLATTKRRHOOTITIPKSSLSVPYVI 900  
 DB 841 RNEQVEIRAVLYNFRQNELKVRVELLHNPAFCSLATTKRRHOOTITIPKSSLSVPYVI 900  
 QY 901 VPLKTLGLOEVEKAAVYHHFISDGVKRSKLVKVPREGIRMNKTAVRVLDPERLGRGVQKE 960  
 DB 901 VPLKTLGLOEVEKAAVYHHFISDGVKRSKLVKVPREGIRMNKTAVRVLDPERLGRGVQKE 960

QY 961 DIPPADLSQVPTSETRILLOCTPVAQMTEDAVDAERLKHLLVTPSGGCEQNMIGTWP 1020  
 DB 961 DIPPADLSQVPTSETRILLOCTPVAQMTEDAVDAERLKHLLVTPSGGCEQNMIGTWP 1020  
 QY 1021 TVIAVHYLDETEQWEKFGLEKQGALELIKKGYTQQLAFRQPSAPAFVKRAPSTWLTA 1080  
 DB 1021 TVIAVHYLDETEQWEKFGLEKQGALELIKKGYTQQLAFRQPSAPAFVKRAPSTWLTA 1080  
 QY 1081 YVVKVESLAVNLAIADSOVLGAVKWLILBKQKPGVFOEDAPVTHOEMIGGLRNNNEKD 1140  
 DB 1081 YVVKVESLAVNLAIADSOVLGAVKWLILBKQKPGVFOEDAPVTHOEMIGGLRNNNEKD 1140  
 QY 1141 MALTAFLVLSLOEAKDICEEQVNSLPFSITKAGDFLEANYMNLQRSYTVIAGVALAQM 1200  
 DB 1141 MALTAFLVLSLOEAKDICEEQVNSLPFSITKAGDFLEANYMNLQRSYTVIAGVALAQM 1200  
 QY 1201 RLKGPLLNKFLITAKDKRWEDPGQLXNVEATSVALLALQLKDFDVPVVRWLNNEOR 1260  
 DB 1201 RLKGPLLNKFLITAKDKRWEDPGQLXNVEATSVALLALQLKDFDVPVVRWLNNEOR 1260  
 QY 1261 YGGGYGSTQATFMVFOALAQYQKDAPOHBLNLDVSLQPSRSKITHRIHWEASLLR 1320  
 DB 1261 YGGGYGSTQATFMVFOALAQYQKDAPOHBLNLDVSLQPSRSKITHRIHWEASLLR 1320  
 QY 1321 SEETKENEGFTVTAEKGGQGTLSVVVTMYHAKADQLTCNKFDLKVITIKPAPETEKRPQDA 1380  
 DB 1321 SEETKENEGFTVTAEKGGQGTLSVVVTMYHAKADQLTCNKFDLKVITIKPAPETEKRPQDA 1380  
 QY 1381 KNTMILEICTRVRGODATMSILDISMNTGAPAPDTDDLKQLANGVDRIYSKYELDKAFSD 1440  
 DB 1381 KNTMILEICTRVRGODATMSILDISMNTGAPAPDTDDLKQLANGVDRIYSKYELDKAFSD 1440  
 QY 1441 RNTLIIYLDKVSHESDDCCLAFKVKHGYFNVELIQGAVKVAAYYNLESCTRFVHPEKEDG 1500  
 DB 1441 RNTLIIYLDKVSHESDDCCLAFKVKHGYFNVELIQGAVKVAAYYNLESCTRFVHPEKEDG 1500  
 QY 1501 KLKLCRDLRCABENCFIQSDDKVTLEBLDKACEPGVDVYVYKTRLVKQVLSNDFDE 1560  
 DB 1501 KLKLCRDLRCABENCFIQSDDKVTLEBLDKACEPGVDVYVYKTRLVKQVLSNDFDE 1560  
 QY 1561 YIMAEQTIKSGSDVQVQQTFTFISPIKCREALKLEEKHYLMWGLSSDFWGEKPNLSY 1620  
 DB 1561 YIMAEQTIKSGSDVQVQQTFTFISPIKCREALKLEEKHYLMWGLSSDFWGEKPNLSY 1620  
 QY 1621 IIGKDTWVHPEDEECODEENKQKQDGLGAFTESMVVFVGCNP 1663  
 DB 1621 IIGKDTWVHPEDEECODEENKQKQDGLGAFTESMVVFVGCNP 1663

## RESULT 3

US-09-142-334-22  
 ; Sequence 22, Application US/09142334  
 ; Patent No. 6268485  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Parries, Timothy C.  
 ; APPLICANT: Harrison, Richard A.  
 ; TITLE OF INVENTION: Down-Regulation Resistant C3 Convertase  
 ; FILE REFERENCE: 4-30443/A/IMU/PCT  
 ; CURRENT APPLICATION NUMBER: US/09/142,334  
 ; CURRENT FILING DATE: 1999-04-15  
 ; EARLIER APPLICATION NUMBER: PCT/CB97/00603  
 ; EARLIER FILING DATE: 1997-03-04  
 ; NUMBER OF SEQ ID NOS: 35  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 22  
 ; LENGTH: 1663  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-142-334-22

Query Match 100.0%; Score 8609; DB 3; Length 1663;  
 Best Local Similarity 100.0%; Pred. No. 0;



Matches 1663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	MGPTGSPSLLLLLLTHPLALGSPMYSIITPNILRESEETMVLSEADHDAQDQVPTVTVH	60
DB	1	MGPTGSPSLLLLLLTHPLALGSPMYSIITPNILRESEETMVLSEADHDAQDQVPTVTVH	60
QY	61	DFPCKKLVLSSEKTVLTPTNHNMGNTFTIPANREFKSEKGNKEVTVQATFGTVVEKV	120
DB	61	DFPCKKLVLSSEKTVLTPTNHNMGNTFTIPANREFKSEKGNKEVTVQATFGTVVEKV	120
QY	121	VLVLSQGYLFIQDXTIYTPGSTVLYRIFTVNHKLLPVGRVMVNIENPEGIPVKQDSL	180
DB	121	VLVLSQGYLFIQDXTIYTPGSTVLYRIFTVNHKLLPVGRVMVNIENPEGIPVKQDSL	180
QY	181	SSQNQLGVLPLSWDIPELVNMGWKIRAYYENSPOQVSTFTEPEVKEVLPSEFIVBTE	240
DB	181	SSQNQLGVLPLSWDIPELVNMGWKIRAYYENSPOQVSTFTEPEVKEVLPSEFIVBTE	240
QY	241	KFYIYINKEGLEVTITARELYGKKVEGTAFVFGIQDGEORISIPESLKRPIEDGSGEV	300
DB	241	KFYIYINKEGLEVTITARELYGKKVEGTAFVFGIQDGEORISIPESLKRPIEDGSGEV	300
QY	301	VLSEKVLLDGQVNPRAEDLVGKSLVYSATVILHSGSDMVQAERSGIPVTSPIQIHFTKT	360
DB	301	VLSEKVLLDGQVNPRAEDLVGKSLVYSATVILHSGSDMVQAERSGIPVTSPIQIHFTKT	360
QY	361	PKYFKPGMPDLNVFTNPDGSPAYRVPVAVOGETVQSLTQDGVAKLSINTHPSOKPL	420
DB	361	PKYFKPGMPDLNVFTNPDGSPAYRVPVAVOGETVQSLTQDGVAKLSINTHPSOKPL	420
QY	421	SITVTRTKQELSEAEQATRTMQALPYSTVGNSENNYLHLSVLTRELPGETLNVNFLRMD	480
DB	421	SITVTRTKQELSEAEQATRTMQALPYSTVGNSENNYLHLSVLTRELPGETLNVNFLRMD	480
QY	481	RAHEAKIRYTYLIMNKGRLKAGROVREPQGLVPLSITTDPIPSFRVAVAYTLGA	540
DB	481	RAHEAKIRYTYLIMNKGRLKAGROVREPQGLVPLSITTDPIPSFRVAVAYTLGA	540
QY	541	SGQREVVDVVDVXDCSVGLSVKSGQSEDRQVPGQMTLKIEGDHGARVVLVAVDK	600
DB	541	SGQREVVDVVDVXDCSVGLSVKSGQSEDRQVPGQMTLKIEGDHGARVVLVAVDK	600
QY	601	GVFLVNLKKNLTOSKTDVVEKADIGCTPGSKDVGAVFSDAGLTFTSSSQCOQAORAE	660
DB	601	GVFLVNLKKNLTOSKTDVVEKADIGCTPGSKDVGAVFSDAGLTFTSSSQCOQAORAE	660
QY	661	QCQOPAAARRRSVQLTEKRMKDKYKYPKELRKCCEGDMRENPMRFSQORRTFRISLGEAC	720
DB	661	QCQOPAAARRRSVQLTEKRMKDKYKYPKELRKCCEGDMRENPMRFSQORRTFRISLGEAC	720
QY	721	KVFLDCCNVTILRRQHPASHGLHARSNLDDEIIAREINIVSRSEPPESLWNVDELKE	780
DB	721	KVFLDCCNVTILRRQHPASHGLHARSNLDDEIIAREINIVSRSEPPESLWNVDELKE	780
QY	781	PKKNGISTKLMNPLKDSITTEILAVSDSKGICVADPPEVTVMQDFFIDLRLPSVY	840
DB	781	PKKNGISTKLMNPLKDSITTEILAVSDSKGICVADPPEVTVMQDFFIDLRLPSVY	840
QY	841	RNEQVEIRAVLYNVRQNELKVRVLELHNPAFCSLATTKRRHQQTITIPPKSSLSVPYVI	900
DB	841	RNEQVEIRAVLYNVRQNELKVRVLELHNPAFCSLATTKRRHQQTITIPPKSSLSVPYVI	900
QY	901	VPLKGTQVEVKAAYVHHFISQVREKSLKVVPEGRMKNKTVAVRTLDPERLREGVQKE	960
DB	901	VPLKGTQVEVKAAYVHHFISQVREKSLKVVPEGRMKNKTVAVRTLDPERLREGVQKE	960
QY	961	DIPPADLSQVPTTESETRILOQTPTVAQMTEDAVDAERLKHILVTPSGCGEQNMIGTWP	1020
DB	961	DIPPADLSQVPTTESETRILOQTPTVAQMTEDAVDAERLKHILVTPSGCGEQNMIGTWP	1020
QY	1021	TVIAVHVLDETEQWEKFGLEKRGALELIKGVTQOLAFROPSSAPAFVRAESTWLT	1080
DB	1021	TVIAVHVLDETEQWEKFGLEKRGALELIKGVTQOLAFROPSSAPAFVRAESTWLT	1080

QY	1081	YVVKVFLAVNLIAISQVLCGAVKWLILEKQKPGVFOEDAPVTHQEMIGLNNNEKD	1140
DB	1081	YVVKVFLAVNLIAISQVLCGAVKWLILEKQKPGVFOEDAPVTHQEMIGLNNNEKD	1140
QY	1141	MALTAFLVLSIQAKDICEBOVNSLPGSITKAGDFLEANYMNLORSYTVATAGVALAQMG	1200
DB	1141	MALTAFLVLSIQAKDICEBOVNSLPGSITKAGDFLEANYMNLORSYTVATAGVALAQMG	1200
QY	1201	RLKGPLANKFLTTAKOKNRWEDPGKOLYNVEATSVALLALLOKDFDVPVFWLNEQR	1260
DB	1201	RLKGPLANKFLTTAKOKNRWEDPGKOLYNVEATSVALLALLOKDFDVPVFWLNEQR	1260
QY	1261	YVGGYGSTOATFMVFOALAQYOKDAPDHOELNLDVSLQPSRSSKITTHIHWSASLLR	1320
DB	1261	YVGGYGSTOATFMVFOALAQYOKDAPDHOELNLDVSLQPSRSSKITTHIHWSASLLR	1320
QY	1321	SEETKEGFTVTAEGKQOGLSVVTHYHAKADQDTCNKFPLKVTIKPAPETERPQDA	1380
DB	1321	SEETKEGFTVTAEGKQOGLSVVTHYHAKADQDTCNKFPLKVTIKPAPETERPQDA	1380
QY	1381	KNTMILEICTRYRGDQDATMSILDISMTGTFAPDITDLDKQLANGVDYRISKYELDKAFSD	1440
DB	1381	KNTMILEICTRYRGDQDATMSILDISMTGTFAPDITDLDKQLANGVDYRISKYELDKAFSD	1440
QY	1441	RNTLIIYLDKVSHEDDCLAFKVOYFNVELIOPGAVKUYAYNLEESCTEFYHPEKEDG	1500
DB	1441	RNTLIIYLDKVSHEDDCLAFKVOYFNVELIOPGAVKUYAYNLEESCTEFYHPEKEDG	1500
QY	1501	KLNLCDRELRCRAEENCFIQKSDDKVTLEERLDKACBPGVDYVYKTRLVQLNSDFDE	1560
DB	1501	KLNLCDRELRCRAEENCFIQKSDDKVTLEERLDKACBPGVDYVYKTRLVQLNSDFDE	1560
QY	1561	YIMAIBOTIKSGDEVQVQOORTFISPIKREALKLEKKHYLMGLSSDFWGEKPNLSY	1620
DB	1561	YIMAIBOTIKSGDEVQVQOORTFISPIKREALKLEKKHYLMGLSSDFWGEKPNLSY	1620
QY	1621	IIGKDTVWEHPBEDECCQBEENKQCODLGAFTESWVFGCPN	1663
DB	1621	IIGKDTVWEHPBEDECCQBEENKQCODLGAFTESWVFGCPN	1663

RESULT 4  
US-08-447-411-2  
Sequence 2, Application US/08447411  
Patent No. 5773243  
GENERAL INFORMATION:  
APPLICANT: FRITZINGER, DAVID C.  
APPLICANT: BREDEHORST, REINHARD  
APPLICANT: VOGEL, CARL-WILHELM  
TITLE OF INVENTION: DNA ENCODING COBRA C3, CVF1, AND CVF2  
NUMBER OF SEQUENCES: 81  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,411  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/043,747  
FILING DATE: 07-APR-1993  
ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5773243man F.  
 REGISTRATION NUMBER: 24,618  
 REFERENCE/DOCKET NUMBER: 1126-101-0  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 413-3000  
 TELEFAX: (703) 413-2220  
 TELEX: 248855 OPAT UR  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1651 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-447-411-2

Query Match 51.0%; Score 4389.5; DB 1; Length 1651;  
 Best Local Similarity 51.3%; Pred. No. 0;  
 Matches 853; Conservative 309; Mismatches 471; Indels 31; Gaps 18;

QY 9 LLLLLTHLPLALGSPMYSIITPNILRLSEETMVLAEHDAOGDVPVTVVTHDFPGKLV 68  
 DB 8 LVAALLGFGSGHAGLYTILITPAVLRTEQILVEAGDSTPKSLDIFVHDFPRKQT 67  
 QY 69 LSSEKTVLTATNMGNVTTIPANKEFKSEKGRNKFTVQATFTQV-VEKVLVLSQS 127  
 DB 68 LFSQSRVDMNAGSMEFTPTTKVPA-KELNKSQNCQVYVVKVT-GPOVALEKVVLLSYQS 125  
 QY 128 GYLPIQTDKIYTPGSTVLRIFTVNHKLLPVGRVTVMNIENPEGIPVKQDSLSSONOLG 187  
 DB 126 GFVFIQTDKIYTPGSTVLRIFTVNHKLLPVGRVTVMNIENPEGIPVKQDSLSSONOLG 184  
 QY 188 VLPLSMDIPELVNGQWIKIRAYENSQQVFSTEFVEKVEYVLPSEFVIVPEKPYIYN 247  
 DB 185 --RPNYLPVLSFGTWKAVAKYEHSPESYATAYDFVREYVLPSEFVIVPEKPYIYN 241  
 QY 248 EKLGVITITARFLYGGKVEGTAFVIGIODGEORISLPSLAKIPIEDSGEVLRSKVL 307  
 DB 242 NKPHVSITARYLYGKVEGTAFVIGIODGEORISLPSLAKIPIEDSGEVLRSKVL 301  
 QY 308 LDGVNPRADLVKSLYVSATVILHSGSDMQAERSGIPVTSYQIHFTKPYKPG 367  
 DB 302 RSREPD--LNQLVGHLYVSVTVITBESGDMVTVTEQGIHIVTSPYQIYFTKPYKPG 359  
 QY 368 MPDPLMVFTNPDGSPAYRPV--AVQGEDTVQSLTQGDGVAKLSINTHPSQKPLSTV 424  
 DB 360 MPYELTVYVNPDGSPAAHVVPVSEAHSEGT---TLSDGTAKILNTPNIQSLPTV 415  
 QY 425 RTKQELSEAEQATRTMQALPYSTVGNNSYHLVLRTELPBGTNLNVLNLRMDRAHE 484  
 DB 416 RTNKGDLPRERQAIKSMATATAYQTQGSSENYLHVATITSEIKPGDNLPLVNFVNRGNANSL 475  
 QY 485 AKIRYTYLIMNKGRLKAGQVREPGQDLVLPISITTDFTPSFRLVAYTYLICAGSOR 544  
 DB 476 NQIKFYIYLINLKGKIFKVGQRPDRDQNLVTLNHLITPDLPSPFRVAYYQV---GN 531  
 QY 545 EVVADSVVDVKSVCVGLVVKSGS-EDRPVFGQOMTLKIEGHDGAVLVAVDKGVF 603  
 DB 532 EVVADSVVDVKTCTMGTLVVGASSRDRRIQKPGAAKIKLEGDPGARVGLVAVDKAVY 591  
 QY 604 VLNKNKLTOSKVDVVEKADIGCTPGSGDKVAGVPSDAGLFTSSSGQQAQRAELQCP 663  
 DB 592 VLNDKYKISQAKIWDITIEKSDFGCTAGSGQNLGVEDAGLALTSTNLNLTQKRAAKCP 651  
 QY 664 QPAARRR--SVQLTEKMDKVGKYPKE-LRKCCEDGMENPNRPFSCQRTFRPISGEACK 721  
 DB 652 QPANRRRRSVLLLDKSKASAAQFQDQGLKCCEDGMENPNRPFSCQRTFRPISGEACK 711  
 QY 722 KVFLLDCNYYTELRRQARASHLGLARSNLDIIAEENIVSRSEFPESLWNNVEDLK-E 780  
 DB 712 AAFLECCYHKGIEDENCRSELSFLARSDFEDELFGDDNIIERSDFPESLWNLTELTGE 771  
 QY 781 PPKNGISTKLMNIFLKDSTITWAILAVMSDKKIGCVADPFVTVMQDFFIDLRPLPSV 840

DB 772 PNNQGISSKTVPFYLRDSITTWELLAVGLSPTKGIQVAPYETIWMKDDFFIDLRPLPSV 831  
 QY 841 RNEQVEIRAVLYNRYQONELKVRVELLHNPAFCSLATTKRRHQOTITIPPKSSLSVPIV 900  
 DB 832 KNEQVEIRAILYNY-ADEDIYVRVELIYNPAFCASASTEGORYQOFPKALSSRAVPFI 890  
 QY 901 VPLKTLQGEVEVKAAYVYHFIISDGVKSLKVVPEGIRMNKTVAVRTLDPERLREGVQKE 960  
 DB 891 VPLEQGLHDEVEIASVRGELASDGVKRLKVVPEGERKN-IVTIELDPSVKGVGQTEL 949  
 QY 961 DIPPADLSQVDPDTESETRILLQGPVPAQMTEDADVAERLKLHIVTSPGCGEONMIGWTP 1020  
 DB 950 TVIANKLDDKVPDTEVETRIISVLGDPVAQIENSIDGSKLNLHITSPGCGEONMIGWTP 1009  
 QY 1021 TVIAVHLDTEQWKEFKLEKQGALELIKKGYTOQLAFROPSSAPAFVRAKRASTWLT 1080  
 DB 1010 SVIATYLDATGQWENLGVDRTEAIKQIMTGYAQOMVYKKAHSAFTNRASSWLT 1069  
 QY 1081 YVVKVPSLAVNLIA-IDSOVLGAVKWLILEKQKPDGVQFQEDAPVIHOEMIGGLRNNEK 1139  
 DB 1070 YVVKVLAASNNVKKDISHEIICGVKWLILNRPQDGVFKENAPVIHGEMLGGTK-CAEP 1128  
 QY 1140 DMALTAFLVLSIOEAKDICEQVNSLPGSITKAGDFLEANTMNLORSYTVIAIAGYALAQM 1199  
 DB 1129 EASLTAFIVTALLERSVCKEQINTLDSINKATDYLKKYEKLQRPYTTALTAYALAAA 1188  
 QY 1200 GRLLKGPLLNKFLTTAKDKNRWEDPGKLYNVEATSYALLALLQLKDPDFVPPVVRWLNBEQ 1259  
 DB 1189 DLANDDEV--LMAASTGRNWEENYARNTHNIEGTSYALLALLKMKKPAEVPVVRWLDQ 1246  
 QY 1260 RYGGGYGTOATFVFOALAQYKXADPHQELMNLDSVLQPSRSSKITHRIHWESALL 1319  
 DB 1247 KYGGTYGTOATFVFOALAEYEIQMPTHQDLNLDISIKLPEREVPERYSINDRNAVQA 1306  
 QY 1320 RSEETKENGSGFTVTAEGKGGTSLSVVVMYHAKAK-DQLTCKNPKDLKVTIKPAPETERKPQ 1378  
 DB 1307 RTVETKLNEDFTVSASGDKATMTILTVYNAQLREDANVCNPKHLDVSVENVELNLQAK 1366  
 QY 1379 DAKNWMILBEITRYRGDQDATMSILDISMTGFPAPDPTDQLKLANGVDVRYISKYELDKAF 1438  
 DB 1367 GGKAAURLKICRYLGEVDSTMTIISMLTGFFPDAEDLKLGLNGVDVRYISKYELDNNM 1426  
 QY 1439 SDANTLIIYLDKSHSEDDCLAFKVOYFNVLELQPGAVKYVYVYNNLEESCTRYVHPEKE 1498  
 DB 1427 AQGTWVYLDKSHSEDECLHFKHKEFVGFTQPGSVKYVYNNLEDEQCTKRYHPEKE 1486  
 QY 1499 DGKMLKCRDELRCBAENCFIQKSDSKVTLEERLDKACEFGVDYVYKTRLVKQLSNDP 1558  
 DB 1487 TGLNKKICHGNTCRCAEETCSLLNQKKIDILQRIQKACAQNVYVYVYKTRLVKQLSNDP 1546  
 QY 1559 DEVYMAEQTIKSGSDEVQVQGRFTFISPIKCRALKLEEKHYLMWGLSSDFWGEKPNL 1618  
 DB 1547 DIVFMDVLEVIKGGTDRNAQAQARQYVSRKQCEALNKLNDVYLINGLSSDLWPMKDDI 1606  
 QY 1619 SYIIGKDTWVEHWPEDEQDEENOKQCDQLGAFTESMWVFGCP 1662  
 DB 1607 SYLITKNTWIERPNWDEQDEEFQNLCDLDFQAQSLNTLTIFGCP 1650

## RESULT 5

US-08-662-227-2  
 ; Sequence 2, Application US/08662227  
 ; Patent No. 5922320  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VOGEL, CARL-WILHELM  
 ; APPLICANT: BREDEHORST, REINHORST  
 ; APPLICANT: KOCK, MICHAEL  
 ; APPLICANT: FRITZINGER, DAVID  
 ; TITLE OF INVENTION: RECOMBINANT PROCVF  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
 ; ADDRESSEE: P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/662,227  
FILING DATE: 14-JUN-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 1126-0107-0X  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1642 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-662-227-2

Query Match 49.1%; Score 4224; DB 2; Length 1642;  
Best Local Similarity 49.6%; Pred. No. 0;  
Matches 824; Conservative 324; Mismatches 478; Indels 36; Gaps 20;

Qy	9	LLLLLLTHPLALGSPMYIIIPNLRLESEBTMLRHAHQGDPVTVVHDPFGKIV	68
Db	8	LVAALLIGPGSHGALYTLTPAVLRDTEBQIILVEAHGSDTPKQDIDFVHDPFKQT	67
Qy	69	LSSEKTVLPATNMGNVFTTPANREFKSEGRNKFVTVQATFGTV-VEKVLVLSQS	127
Db	68	LFQTRVDMNPAGGLMVTPIEIPA-KEVSTDQRQNVVVQVTP-GPQVRLEKVVLSYSQS	125
Qy	128	GYLFQTDKTIITPGSTVLRYRFTVNHKLLPVGRWTWNIENPGIPVKQDSSLSSQNLG	187
Db	126	SFLFQTDKGIYTPGSPVLRYRFSMDHTSNKNKTVIVEFQTPGILVSSNSVD-----	179
Qy	188	VLPLSW--DIFELVNMGWKIRAYYENSPQVSTEFSEVKEBYVLPSPSEVIEPTKEFYI	245
Db	180	-LNFPPYLPDLVSLGTWRIIVAKYHSPEN-YTAYFDVRKYVLPSPSEVRLQPSKEFFYI	237
Qy	246	YNEKGLVITAFPLGKVEGTAFVIGIQDGEORISLPSLXRIPIEDSGSEVLSRK	305
Db	238	DGNENFVSLTAKILYGESEVGFVLFQKIDDAKSIPLSTRIPIDIDGDKATLKD	297
Qy	306	VLDGQNPRAEDLVGKSLYSATVILHSGSDMVQASRGIPVTSYQIHFTKPYFK	365
Db	298	TFRS--RFNMLNVLGHTLYASVTVMTSGSDMVVTEQSGIHIIVASPYQIHFTKPYFK	355
Qy	366	PGMPDLMVFTNPDGSPAYRVPVAVOGEDTVQSITQDGVAKLSINTHPSQPLSITVR	425
Db	356	PGMPYELTVFTNPDGSPAAHVPVSEAFHSN-GTLLSDGFAKILNPLNAQSLPTIVR	414
Qy	426	TKSQELSEAEQATRTMOALPYSTGVNSNNYLSLVLRTLPBGTLLNVLNLLRMDRAHEA	485
Db	415	THRGDLFRERQATKSMATAYQTQGGSGNYLHVAITSTEIKFGDNLNPNFNVKNANSLK	474
Qy	486	KIRYTYLLMNGKLLKAGQVREPQDLVLPISITTDFTPSRELVAITYLLIGASQRE	545
Db	475	QIKYFYLLNKGKIFKVGQRPRDQGNLVTWNLHITPDLIPSPFRFVAYVQV-----GNNE	530
Qy	546	VVADSVVVDVKS CVGSLVVKSGQSDRPQVPGQMTLKIEGDHGARVVLVAVDKGVFL	605
Db	531	IVADSVVVDVKOTCMGTLVVK--GDNLTQMEGAAMKIKLEGDFGARVGLVAVDKAVYVL	587

Qy	606	NKQKLTQSKIMDVVKADIGCTPGSGKQYAGVFS DAGLTFTSSSQOQTAQBAELQCPQ	665
Db	588	NDKYISQAKIMDTIEKSDFGCTAGSGQNNLGVFEDAGLALTSTNLTQKSAACQCPQ	647
Qy	666	AARRRR-SVOLTEKRMQKVGKY-PKELRKCCEGDMRENPMRFSCQRRTRFISLGSACKKV	723
Db	648	ANRRRRSVLLDSNASKAAAFQDQLRKCCEDWHENPMGYTCBKRAKVIQEGDACKAA	707
Qy	724	FLDCCNVITELRQHARASHLGLARSLDDEIIABENIVSRSEFESMLWNVEDI-KEPP	782
Db	708	FLECCRYIKGVDRDENQRESELFARDNEDGFIADSDIISRSDFKASMLWLTQDTEPN	767
Qy	783	KNGISTKLNNIFLKDSITTWELLAVSMDDKIGICVADPEFVTWQDFFIDELPYSVVRN	842
Db	768	SGIISKTMGTYLRDSITTWVLAWSFTPTKIGICVAEPYRVMKVFFIDLOMPYSVVRN	827
Qy	843	FOVEIRAVLYNRYONQELKYRVELLNPAFCSLATTKRHHQOTIIPPKSSISLVPYIVP	902
Db	828	EQVEIRAILLNY-VNEDIYRVELLNPAFCSSASTKGRIYRQOPFIKALSSRAVFEVIVP	886
Qy	903	LKTLQEVVEKAAVYHHFISDGVKSLKYVPEGIRMNKTVAVRTLDPERLGEVQKEDI	962
Db	887	LEQGLHDEVEIKASVOEALMSDGVKSKLVVPEGVQ-KSIVTVIKLDPRAKGVGGTQLEVI	945
Qy	963	PPADLSQVDPDTESETRILLQGTPEAQMTEDAVDAERLAKHLIWPSSGCGEQNMIGMTFTV	1022
Db	946	KARKLDRVPDTELETKIILQGPVPAQITENSIDGSKLNHLIITPSGCGEQNMIMAAPV	1005
Qy	1023	IAYHLDTEQWKEFGLEKQGALEIHKGYTQQLAFROPSSAFAPVFKRPASTWLTAYV	1082
Db	1006	IATYLTDTTEQWETLGNRRTEAVNQIVTGYAQOMVYKKAADSHSYAFTNRASSWLTAYV	1065
Qy	1083	VKVESLAVNLIA-IDSOVLGAVKVLLEKOKPDGVQEDAPVHQEMIGGLRNNEKDM	1141
Db	1066	VKVFAAKVAVGSHIEICGVFWILLNRQDQDGAFAKNAFVLSGTWGGIQC-CAEEV	1124
Qy	1142	ALTAPVILISQAKDICEEOVNSLPSITKAGDFLEANTMNIORSYTVVAIAGYALAQMR	1201
Db	1125	YLTAFILVALLESKTIENDVNSLDSISKATNYLLKYEKLQRPYTTALTAYALAAADQ	1184
Qy	1202	LKGPLANKFLTAKDKNRWEDPGKQYNYVEATSYALLALLQLKDFDPVPPVVRWLEORY	1261
Db	1185	LNDRVLMASSTGRD--HNEEYNAHTNTEGTSYALLALLKMKKDFDQGPVFWLTQNP	1242
Qy	1262	YGGYGTQATFWVFOALAQYKDAFDHQLNLDVSLQPLPSRSKITHRIHWEASLRS	1321
Db	1243	YGETYGTQATWAFQALAEYEIQMPTKDLNLDITIELDFREVPIRYRYNTENALLART	1302
Qy	1322	EETXENSGFTVTABGKGGTSLVVVTVYHAKAKDOL-TCKNFDLKVTKIPAPETEKRPDA	1380
Db	1303	VELKLNODITVTSAGDGAATWLTITFNAQLQEKANVCNKFHLNVSVE---NIHLNMA	1359
Qy	1381	KNTWILICITRYRQDQATMSILDI SMGTGFAPTDLDLQLANGVDVRYISKTTELKAFSD	1440
Db	1360	KGALMLKICITRYLGEVDSTWTIIDISMLTGLFPAEDLTRLKSGVDVRYISRYEYDNNMAQ	1419
Qy	1441	RNTLIIILDKVSHSEDDCLAFKHQVFNVELIQFGAVKVYVAYNLEESCTRYHHPKEDG	1500
Db	1420	KVAVIIVLNKSHSEDECLHFKILKHFEVGFQPGSVKYSYNNLDEKCTKYHDPKGTG	1479
Qy	1501	KLNLCRDELCRCABENCFIQKSDSKVLEERLDKACBPGVDYVYKTRVLKQLNSDFE	1560
Db	1480	LLNKICIGNVCRCAGETCSLNLQERIDVPLQIEKACETNVDYVYKTRVLKRLIEEQDGNDI	1539
Qy	1561	YMAIEOTIKSGSDEVQVQOORTFIPGICREALKLEKKHYLMWGLSDFDGEKPNLSY	1620
Db	1540	YVMDVLSVVKQDTEENPRAKTHQYISQKQOEALNKVNDVYILWGSRSDDLPTKDKISY	1599
Qy	1621	IIGKDTVWEHWPDEDEQDDEENQXQDGLGAFTESEVWFGCP	1662
Db	1600	IITKNTIWRPHDEDECEBEFQKLCDDDDFAQFSYTLTEFGCP	1641





QY 1142 ALTAFLVLSIQEAKDICEQVNSLPGSITKAGDFLEANYVNLQSRSTVVAIGVALAQMGR 1201  
 Db 1145 YLTAFLVALLSSTKNDYNSLSSIKATNLLKXKYLQRPYTTALTAYALAAADQ 1184  
 QY 1202 LKGPLLNKFLTAKDKNRWEDFGKQXNVEATSVALALLQLKDFDFVPPVVRWLNQRY 1261  
 Db 1185 LNDRLVMAASTGRD--HWEYNATHNIEGTSYALLAKMKKFDQTPVIEWLTQNF 1242  
 QY 1262 YGGVGSQTAFMYFOALAQKQDAPDQHSINLDSLOLPSRSSKTHRIHESASLLRS 1321  
 Db 1243 YGETYGGQATMAFQALAEYIQMPTHKOLNLDITIELDPREVPIRYINVENALLART 1302  
 QY 1322 PETRENEGFTVTAEGKGGOTLSVVMTYHAKAKDQL--TCNKFDLKVTKIKAPETERQDPA 1380  
 Db 1303 VETKLNQDITVTSAGDGKATMTILTFFNAQLQKQKANYCNKFLNVSVE--NIHLNAMGA 1359  
 QY 1381 KNTMLEICTRYRGDQATWILDSIMMTGFAPDQDQLKQANGVDVRYISKYELDKAFSD 1440  
 Db 1360 KGALMLKCTRYLGEVDSMTIIDISLMTGLPDAEDLTRLSKGVDRYISRYEVONNMAQ 1419  
 QY 1441 RNTLIILDKYSHSEDDCLAFKHQYFNVELIQPGAVKVVAYYNLEESCTRFYHPEKEDG 1500  
 Db 1420 KVAVIILNKYSHSEDECLHFKILKHEVGFQPGSVKVSYYNLDEKCTKPYHDKGTG 1479  
 QY 1501 KLNKLCRDELRCRAEENCFOKSDDKYTLBERLDKACEPGVDVYKYLKVKQLSNDPDE 1560  
 Db 1480 LNKICIGNVCRAGETCSSLNHQRIDVPLQEKACETNVDVYKYLKILRIEEQGNDI 1539  
 QY 1561 YMAIBQTKSGSDEVQVQOQRTFISPIKCREALKLEEKHYLMWGLSDFWGEKPNLSY 1620  
 Db 1540 YVMDVLEVIKQGTDENFRKTHQVYSQKQCEALNKLKNDYILWGSRLDLPKDKISY 1599  
 QY 1621 IIGKDTWVHWPEDDECODEENOKQCODLGAFTESMVVRGCP 1662  
 Db 1600 IITKNWIERPHEDECOEEBFQKLCDDFAQFSYLLTFEGCP 1641

## RESULT 8

US-08-447-411-45  
 ; Sequence 45, Application US/08447411  
 ; Patent No. 5773243  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FRITZINGER, DAVID C.  
 ; APPLICANT: BREDEHORST, REINHARD  
 ; APPLICANT: VOGEL, CARL-WILHELM  
 ; TITLE OF INVENTION: DNA ENCODING COBRA C3, CVF1, AND CVF2  
 ; NUMBER OF SEQUENCES: 81  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/447,411  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/043,747  
 ; FILING DATE: 07-APR-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Oblon, No. 5773243man F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 1126-101-0  
 ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 413-3000  
 ; TELEFAX: (703) 413-2220  
 ; TELEX: 248855 OPAT UR  
 ; INFORMATION FOR SEQ ID NO: 45:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1642 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-447-411-45

Query Match 48.9%; Score 4208; DB 1; Length 1642;  
 Best Local Similarity 49.5%; Pred. No. 0;  
 Matches 822; Conservative 322; Mismatches 482; Indels 36; Gaps 20;

QY 9 LLLLLLTHLPLALGSPWYSIITPNILRLSEETMVLAEHAQDQDPVTVTVHSDRPGKLV 68  
 Db 8 LVAALLIGFPGSGHGLYLIITPAVLRDTQTEQILVEAHGDSPTKQLDIFVHDFPRKQT 67  
 QY 69 LSSEKTVLTATNMGNVTFTIPANREFKSEKRNKFTVQATFGTV--VEKVLIVLSQS 127  
 Db 68 LFQTRVDMNPAGGMLVTPTEIPA--KEVSTDSRQNYVVVQVT--GFQVRLKXVLLSYQS 125  
 QY 128 GYLFIQTDKTIYTPGSGTVLYRIFTVNHKLLPVGRVTVMNIENPEGIPVKQDSLSSQNQLG 187  
 Db 126 SFIFIQTDKGLIYTPGSPVLYRVFSDMHNSTKMKVIVIEFQTPGEILVSSNSVD----- 179  
 QY 188 VLPLSW--DPELVNNGQWKIRAYENSPQQVFSEFEVKEYLVLSFSEFVEPTKFPYI 245  
 Db 180 -LNFPPYNPDLVSLGTWRIVAKYHSPEN--YTAFFDVVKYVLPSEFVRLQPSSEKFFYI 237  
 QY 246 YNKGLEVTITAFPLGKXVEGTAFFVIGIQDGEORISLPESLRIPEDTSGSEGVVLARK 305  
 Db 238 DGNENFHSITARYLYGEEVGVAFVFGVKIDDKAKSIDPSLIRIPIDGGKATLRD 297  
 QY 306 VLLDGVQNPRAEDLVCKSLYVSATVILHSGSDMWQAERSGIPVTSYQIHFTKTPKPYK 365  
 Db 298 TFRS--RFPNLNVLGHVTLVASTVMTESGSDMWVTEQSGIHIVASPYQIHFTKTPKPYK 355  
 QY 366 PGMPFDLMVFTNPDGSPAVRVPVAVQGEDTVQSLTQGGVAKLSINTHSQKPLSITVR 425  
 Db 356 PGMPYELTVYVTPDGSPAAHVPVSEAFSH--GTTLSGDTAKLILNPLNAQSLPIIVR 414  
 QY 426 TKKQELSEABQATRTMQALPYSTVGNSSNNYHLVRLTELREGTELNVNPLLRMDRAHEA 485  
 Db 415 TNEGDLPRERQATKSMATAAYQTQGGSGNYLHVATISTEIKEDGNLPVKFQCEGQNSLK 474  
 QY 486 KIRVYTVLIMNKGELKAGQVREPGQDLVVLPLSITTDFFPSRPLVAYYTLIGASGORE 545  
 Db 475 QIKFYVILLINKKIFKVGQRPRDQCNLVNWLHITPDLPSPFRFVAYQV----GNNE 530  
 QY 546 VVADSVWVDVKSCVGLSVVKGQSEDRQPVPGQOMTLKIEGDHGAARVVLVAVDKGVFL 605  
 Db 531 IVADSVWVDVKDTCMGTLVVK--GDNLIQMPGAAMKIKLEGDPGARVGLVAVDKAVTVL 587  
 QY 606 NKNKLTQSKIWDVVEKADIGCTPGSGKDYAGVSDAGITFTSSSQGQTAQRAELQCPQP 665  
 Db 588 NDKYKISQAKINDTIEKSPGCTAGSQNNLGVFEDAGLALTSTNLNLTKRSAAKCPQP 647  
 QY 666 AARRR--SVOLTEKRMKVGKY--PKELRKCCBDGRENPMRPSQORRTFRFISLGRACKV 723  
 Db 648 ANRRRSSVLLDSNASKAAEFQDQDLRKCCEVMHENPMGYTCERAKYIQEGDACKAA 707  
 QY 724 FLDCCNVITELRRHARASHLGLARSLNLDIEDIAENIVSRSEFFESMLWNVDEL--KEPP 782  
 Db 708 FLECCRYIKGVRENQRESELEFLARDNEDGFIADSDIISRSDFPKSLWLTKLTTEPN 767  
 QY 783 KNGISTKLNIPLKDSITTWELAVMSMDKIGCVADPFVTVMDFFIDRLRLPYSVVRN 842  
 Db 768 SQGISSTKMSFVLKDSITTVVLAISVFTPTKGCICVAEPYEVIRVMKVFFIDLPMPYVVK 827  
 QY 843 EQVEIRAVLYNRQNOELKVRVELLHNPAFCSLATTKRHQOTITIPKFSLSLSPYVIVP 902

Db 828 EQVEIRAILHNY-VNEDIYVRVELLYNPAFCASASTKQRYRQCFPIKALSSRAVPFVIVP 886  
 QY 903 LKTGLQVEVEKAAVYHFFISDGRKSLKVVPEGIRNMKTAVRTLDPERLREGVQKEDI 962  
 Db 887 LEQGLHDEIKASVQKALSDGVRKXKLVVPEGVQ-KSIVITVKLDPRKAVGGTQLEVI 945  
 QY 963 PPADLSQVPDDESETRILLOTPVAQMTEDAVDAERLKHIVTPSGGGEONMIGMPTV 1022  
 Db 946 KAKLDRVPDTEIEFKIIIOGDPVAQIENSIGSKLNHLIITPSGGGEONMIRMAFPV 1005  
 QY 1023 IAVHYLDETEQWKEFKLEKRRQGALELIKGYTQCLAFRQPSAFAAFVKRAPSTWLTAYV 1082  
 Db 1006 IATYLLDTTEQWETLGINRTEAVNQIVGYAQQVMVKKADHSYAAFTNRASSSMLTAYV 1065  
 QY 1083 VKVPSLAVNLIA-IDSOVLGGAKVLLLEKQKEDGVFOEDAPVTHQEMIGGLRNNKDM 1141  
 Db 1066 VKVFAAAKMWAGISHIICGGVRWLLNRQOPDGAFFENAPVLSGTWGGIO-GABEEV 1124  
 QY 1142 ALTAFLVLSIQEAKDICERQVNSLPGSIITKAGDFLEANNMLORSYTVAIAGYALAAQVR 1201  
 Db 1125 YLTAFLVALLESKTCINDVNSLSSIKKATNWLKKYKLRPYTTALTAYALAAQ 1184  
 QY 1202 LKQPLINKFLTAKKRWEDFGKQLYNVYATSYALLALLQKDPDFVPPVVRVNLQRY 1261  
 Db 1185 LNDRLVMAASTGRD--HWEENYNAHTNIEGTSYALLALLKWKPKDQGPVIRWLTQNF 1242  
 QY 1262 YGGYGSTQATFMVFOALACYQKADPHQELNLDVSLQLPERSKITHRIHWSASLLRS 1321  
 Db 1243 YGETYGTQATVMAFQALAEYEQMPTHKDLNLDITIELDREVPVIRVINYENALLART 1302  
 QY 1322 BEYKENEPTVTAEGGQGLSVVYMYHAKAKQOL-TGNKPELKVITKPAETETEKRPQDA 1380  
 Db 1303 VETKLNODIIVTASGDKATMTILTFYNAQLQKCANVCNKKHLNVSVS---NIHLNANGA 1359  
 QY 1381 KNTMILEICTRYRGDQDATMSILDI SMMTGFAPDITDLKQLANGVDYRISKYELDKAFSD 1440  
 Db 1360 KGALMLKICRYLGEVDSMTWITIDISMLTGFPLDAEDLTRLSKGVDRVYSRYEVDNNMAQ 1419  
 QY 1441 RNTLIIILDRVSHSEDDCLAFKHQVFNVELQPGAKVYAYNLEESCTRYFHEKEDG 1500  
 Db 1420 KVAIIVYLNKVSSEDECLHFKLHFVGFQPGSVKYSYINDKCTRYFHPDKGTG 1479  
 QY 1501 KLKLCRDELCRAENCFIQSDDKVLTLEERLDKACBPGVDYVYKTLVKVQLSNDFDE 1560  
 Db 1480 LLNKICIGNVCRCAGETCSLHQRIDVPLQIERKACETNDYVYKTLKLRIRIEQDQNDI 1539  
 QY 1561 YIWAIEQTIKSGSDEVQVQQRFTFISPIKREALKLEKKHYLMKGLSSDFWGEKNLSY 1620  
 Db 1540 YVMDVLEVIKQGDKAPRAKTHQYIISQRKQCALMLKVNDDYILWGSRSDDLPTKDKISY 1599  
 QY 1621 IIGKDTVWEHPDEDCQDENQKQCDLIGATGESMWVFGCP 1662  
 Db 1600 IITKNTWIERWHPDEDCQDEEFQKLCDDPAQSYTLTFGCP 1641

## RESULT 9

; Sequence 35, Application US/08662227  
 ; Patent No. 5922320  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VOGEL, CARL-WILHELM  
 ; APPLICANT: BREDEHORST, REINHORST  
 ; APPLICANT: KOCK, MICHAEL  
 ; APPLICANT: PRITZINGER, DAVID  
 ; TITLE OF INVENTION: RECOMBINANT PROCVP  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY  
 ; CITY: ARLINGTON  
 ; STATE: VA  
 ; COUNTRY: USA

; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA: US/08/662,227  
 ; APPLICATION NUMBER: 14-JUN-1996  
 ; FILING DATE: 14-JUN-1996  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: OBLON, NORMAN F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 1126-0107-0X  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-413-3000  
 ; TELEFAX: 703-413-2220  
 ; INFORMATION FOR SEQ ID NO: 35:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1648 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-662-227-35

Query Match 48.7%; Score 4194; DB 2; Length 1648;  
 Best Local Similarity 49.4%; Pred. No. 0;  
 Matches 821; Conservative 324; Mismatches 481; Indels 36; Gaps 20;  
 QY 9 LLLLLLTHLPALGSPVSIITENILRESEETWLEAHDAQGDVPTVTVTHDFPGKLV 68  
 Db 8 LVALLIGFPOSSGALYITITPAVLRTDEEILVEAHGDSPPKQLDIFVHDFPRKQT 67  
 QY 69 JSSEKTVLTATNMGVNTFTIPANREFKSEKGRNKFVTVQATFGTV-VEKVLVSLOS 127  
 Db 68 LQTRVDNPNAGGMLVPTTIEIPA-KEVSTDSQONQYVVQVT-GPQVLEKVVLLSYOS 125  
 QY 128 GYLFTQDTKTYTTPGSTVLYRIFTVNHKLLPVGRKTVYVNIENPEGIPVKQDSILSSQNLG 187  
 Db 126 SFLFLQTDKGYITGSPVLYRVFSMDHHTSKMNKTVIVEFQTEGLVSSNSVD----- 179  
 QY 188 VLPLSW--DIPELVNMCGWKIRAYENSPOQVSTEFVEKVLVPSFVIVPEFTEKFFYI 245  
 Db 180 -LNFPPNPLDPLVSLGTWRIYAKYEHSPEN-YTAVFDVAKYVLPSEFVRLQPSKFFYI 237  
 QY 246 YNEKLEVTITATFLYKGVKVEGTAFVIFGIDGCEQISLPESLKRIPIEDGSGEVVLSRK 305  
 Db 238 DGNENFHSIITARIYLGEEVGAFLVFGVKIDDAKSIPTDSLTRPIIDGDKATLKR 297  
 QY 306 VLLDGVQNPRAEDLVGKSLYVSATVILHSGSDMVQAPRSQIPVTSPIQIHFKTKYFK 365  
 Db 298 IFRS--RFPNLELVGHTLYASVTWMTESGDMVVTESQGIHIVASPYQIHFTKTKYFK 355  
 QY 366 PCMPFDLMVFTNPDGSPAYRVPVAVQGEDTVQSLTQCGDGVAKLSINTIPSOXPLSITVR 425  
 Db 356 PCMPYELVYTNPDGSPAAHVPVVSAPHSM-GTLLSDGTAKLILNIPNAQSLPITVR 414  
 QY 426 TKKQLSEAEQATRTMQALPYSTVGNNSNVLHLSVLTETLRPGETLNVPLFLMDRAHEA 485  
 Db 415 TNHGDLPRERQATKSMATAIAYQTQGGSGNLYLHVATSTETKPGDNLFPVNFVKNANSUK 474  
 QY 486 KIRVYVYLLMKNKGRLLKAGROVREPQDVLVPLISITDTPDFRFLVYVYVITLIGASQRE 545  
 Db 475 QIKYFTYLLLNKGIKFKVGRQPRDGGQNLVTMNLHITPDLI PSRFVAYVQV---GNNE 530  
 QY 546 VVADSVYVDVQDSVGLSVLVKSGSEDRQPVPGQOMLKEGDEGARVVLVAVDKGVFL 605  
 Db 531 IVADSVYVDVQDSVGLSVLVKSGSEDRQPVPGQOMLKEGDEGARVVLVAVDKGVFL 587  
 QY 606 NKKNKLTQSKIWDVVEKADIGCTPGSGKDYAGVPSDAGLFTFTSSSQOQTAQRAELQCPQ 665  
 Db 588 NDKYKISQAKIWDITKSDFCCTAGSQNNLGVFEDAGLALTITSTNLNTKQSAKCPQ 647



```

QY 666 AARRRR-SVOLTEKEMDKVGY-PKELRKCCEGRENEMRPFQCRTRFISIGACKV 723
Db 648 AARRRSSVLLDSNASAAAFQODLRKCCEDVWHENPMGYTCERAKYIQEGDCKAA 707
QY 724 FLDCNNYITELRQOHARSHLGLARSLNDEDIIAENIVSRSEPPESWLMNVDEL-KEPP 782
Db 708 FLECCRYTKGVADENQRESELFLARDNEDGFIADSDIISRSDFPKWMLTKDLTEEPN 767
QY 783 KNGISTKLMNFKDSITTWILAVSMKDKGICVADPFVYVMQDFIDRLRPLYSVVRN 842
Db 768 SGGISGKMSFFLRDSITTWVLAWSFTPTKGCIVABEYRVMKVFIDLOQPSYVKN 827
QY 843 EQVEIRAVLYNRQONELKVRVELLHNPAPGSLATTKRRHQQTITPPKSSLSVPYVIVP 902
Db 828 EQVEIRAILHNY-VNEDIYVVELLYNPAFCASATKQRYRQOPPIKALSSRAVPFVIVP 886
QY 903 LKTLGOEYEVKAAVYHHISGVRSKLVKVPGEIRKMTKAVATLDPERLGRGVQKEDI 962
Db 887 LEQGLHDEIIRKASVQEAALWSGVKRLKVPGEVQ-KSIVTVKLDPRAGVGGTQLEVI 945
QY 963 PPADLSDQVPTSETRILLQSTPVAQMTEDAVDAERLKH1VTPSGCGEQNMIGMTPTV 1022
Db 946 KARKLDRVPDTEIETKIIQGDPAQIIENSIDGSKNLH1ITPSGCGEQNMIRMAAPV 1005
QY 1023 JAVHYLDETEQWEKFGLEKRGALIKKGYTOQLAPQSSAFAPVKEAPSTWATV 1082
Db 1006 IATYVLDTEQWELGINRTEAVNQIVTGAQMVYKKAADHSAFTNRASSSWLATV 1065
QY 1083 KVVFSLAVNLJA-IDSOVLGAVKWLILEKOPGVQFQEDAPVHOEMIGGLRNNNEKDM 1141
Db 1066 KVVFAMAAKAVAGISHETICGVRLIIRQOPDGAFAKENAPVLSGTMOGGIQ-GAEEV 1124
QY 1142 ALTAFLVLSLOEKDICEQVNSLPGSITKAGDPLEANYNLORSYTVATAGVALAQMGR 1201
Db 1125 YLTAFLVALLESKTCINDYNSLSSIKKATNLKLYEKLPQRYTALTATALAADOQ 1184
QY 1202 LKGPLNLKFLTTAKDKRWEDPGKQLVNVETASVALLALQLKDFDFVPPVVRWLNQRY 1261
Db 1185 LNDRLVMAASTGRD--HWESYNATHNIEGTSYALLALLKMKKFDQGTGVISWLTQDNF 1242
QY 1262 YGGVGSSTCATPMVFOALAOQKADPOHOLNLDVSLQLPSRSKITHRHWESASLLRS 1321
Db 1243 YGETYGOQATWNAFQAAYEIQMPTHKDLNLDITIELDPREVPIRYIRINYENALLART 1302
QY 1322 BETKNEGFTVTAEGKOGTSLVVMYHAKAKOOL-TCNKFPLKVTIKPAPETERPQDA 1380
Db 1303 VETKLNQDITVTASGDGRATITLTFTYNAOLQEKANYCNFHLNVSVE--NIHLNAMGA 1359
QY 1381 KNTMLEICTRYRGDQDATTMSILDI-SMWTGAPDQDLDKOLANGVDRIYSKYELDKAFSD 1440
Db 1360 KGMALMKICTRYLGEVDSMTIIDI-SMLTGLPDAEDLTRLSKGVDRYISRYEVDNNMAQ 1419
QY 1441 RNTLIYLDKVSHEDDCLAPKXHQYFNVELIQGAVKYAYNYNLEESCTRFYHPEKEDG 1500
Db 1420 KVAVIYLYNKVSHSEDECLHKILKHEVGFQPGSVKVSYYNLDKCKTKFYHPDKGTG 1479
QY 1501 KNLKCRBLCRCAENCFIQKSDDKYTLERLDDKACEPGVGVYVYKTLVKQLSNDPDE 1560
Db 1480 LNKICIGNVRCAGETCSSLNHQERIDVPLQIEKACETNVDVYVYKTLRLEEQGNDI 1539
QY 1561 YTMAJETQIKSGSDEVQVQORTTISPICKREALKLEKXHYLMWGLSSDFWGEKPNLSY 1620
Db 1540 YVMDVLEVIKQGTDENPRAKTHQVYSQRKQOEALNKLNVDDYLINGSRDLPTDKISY 1599
QY 1621 IIGKDTVWHPEDEDECOENOKOCODLGAFTESMVVFGCP 1662
Db 1600 IITKNTWTERPHDEDECOEEFQKLCDDPAQFSYTLTEFGCP 1641

```

RESULT 10

US-09-017-947-35

; Sequence 35, Application US/09017947

```

; Patent No. 6303754
; GENERAL INFORMATION:
; APPLICANT: VOGEL, CARL-WILHELM
; APPLICANT: BREDEHORST, REINHORST
; APPLICANT: KOCK, MICHAEL
; APPLICANT: FRITZINGER, DAVID
; TITLE OF INVENTION: RECOMBINANT PROCVF
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/017,947
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/662,227
; FILING DATE: 14-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1126-0107-0X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1648 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-017-947-35

```

Query Match 48.7%; Score 4194; DB 3; Length 1648;

Best Local Similarity 49.4%; Pred. No. 0;

Matches 821; Conservative 324; Mismatches 481; Indels 36; Gaps 20;

```

QY 9 LLLLLLTHLPLALGSPMYSIIITENILRLSEETMVLAEHDAQDVPVTVVTHDFPKKLV 68
Db 8 LVAALLGFGSGHGLYLTITPAVLDRTDTESQLVEAHGSDTPKQLDIFVHDFPRKQT 67
QY 69 LSBEKTVLTIPATNMGVNTFTIPANREFPKSEKRNKFTVQATFGTV-VKQVVLVSQS 127
Db 68 LFQTRVDMNPAGMLVTPTEIIPA-KEVSTDSRQNVVVQVT-GPQVLEKVVLLSYQS 125
QY 128 GYLFIQDRTKTYTPGSLVLYRIETVNHKLLPVGRVTVMNIENPEGIPVQKDSLSQNLG 187
Db 126 SFLFIQDRTKTYTPGSLVLYRIETVNHKLLPVGRVTVMNIENPEGIPVQKDSLSQNLG 179
QY 188 VLPLSW--DIPELVNMGNQKIRAYENSPQVSTEFVEYKVLPSFEVIEPTKPYI 245
Db 180 -LNFPPYNPFLDVLSTLWRIVAKYSHSPEN-YTAYFDVRKVLPSFEVIEPTKPYI 237
QY 246 YNEKGLVTTATFLYKKGVEGTAFVFIQDGEORISLPESLKRIPEDSGEVLGRK 305
Db 238 DGNENFVISTARYLYGEEVGVAVLFGVKIDDAKKSIPDSTLRIPIDGGKATLRD 297
QY 306 VLLDGVQNPRAEDLVGKSLYVSATVILHSGSDMVAQERSGIPVTSPIYCIHTKTPKYPK 365
Db 298 TFRS--RFFNLNELVGHITLYASVTVMTEGSDMVAQERSGIPVTSPIYCIHTKTPKYPK 355
QY 366 PGMPFDLWVFVTPNPDGSPAVRVPVAVQGEDTVQSLTQDGVAKLSINTHPSQKPLSITVR 425

```

Db 356 PGMFYELTVVTPDGSAAHVPVSEAFHSM-GTTLSDGTAKJLILNPLNAQSLPIVTR 414  
QY 426 TKQELSEABOARTWALPYSVTGNSNLYLHSLVRLTELRLGETLNVNLLMDRAHEA 485  
Db 415 TNGGDUFREQAATKSMATAYQTQGGSGNYLHVATSTEIKRQDMPLPNFVKNVANSLK 474  
QY 486 KIRYTYLIMKGLLAKGQVREPOODLVPLSLTTFDIPSPFLVAYTYLIGASQORE 545  
Db 475 QIKYFTYLIILNKGIKFKVGRQPRDQONLVTNLNLHTPLDIPSPRFVAYQV---GNNE 530  
QY 546 VVADSVMDVYKSCVGLSVKSSQSDROPVPGQWTLKIEGDHGARVVLAVDKGVFVL 605  
Db 531 IVADSVMDVYKTCMGFLVVK---GNLILQMGAMKIKLEBDFGARVLVAVDKAVYL 587  
QY 606 NKXNKLTQSKIDWVYKADIGCTPGSGDYAGVFSADGLTFTSSSGQQTARAEQLQCPQ 665  
Db 588 NDKYKISQAKIWDITIEKSDFGCTAGSGQNNLGVFSDAGLALTTSTNLNTKORSAAKCPQ 647  
QY 666 AARERR-SVOLTEKMDKVGKY-PKELRKCCDGMENPMFSCORTRFTISLGEACKV 723  
Db 648 ANRRRSVLLLSNASKAABFQDQDLRKCCDVMHNPXGTYCEKRAKYOEGDACKAA 707  
QY 724 FLDCNVTIELRQHAARASHGLIARNSLDEDIIAENIVSRSEPPESMLNVEDL-KEPP 782  
Db 708 FLECCRYIKGVRDENORESEFLARDNEDGFIADSDIISRSDFPKMWLMLTKOLTEBP 767  
QY 783 KNGISTKLMNIFLKDSTITWEILAVSMDSKIGICVADPEVTVWODFIDLRLPYSVVRN 842  
Db 768 SQGISKTMSPYLRDSTITWVLAVSFTTKGICVABPEYIIRVMKVFFIDLCMPYSVVRN 827  
QY 843 EQVEIRAVLYNQONELKYRVELLNPAFCSLATTKRRHQOTITIPKSSLSVPYVIVP 902  
Db 828 EQVEIRAILHNY-VNEDIYVYRVELLYNPAFCSTASTKQRYQRPPIKALSSRAVPFVIVP 886  
QY 903 LKTCLOVEVKAAYVHHFISDGVKSLKVUPPEIRNKTAVVRLDPERLREGVQKEDI 962  
Db 887 LEQGLHVDVEIKASVQEAALMSDGVKXKLVVPEGVQ-KSIVTIVKLDPRAGVGGTQLEVI 945  
QY 963 PPADLSQVPTDSETRILLQGTTPVAMTDEADVABRLKHLIVTPSGGGEQNMGMPTV 1022  
Db 946 KARLDDVRDPTETIETKIITQSDPVAQIENSIDGSKLNHLIITPSGGGEQNMIRMAFP 1005  
QY 1023 IAVHLDTEBQWFKLEKQGALELIIKGYTQOLAFROPSSAFKVPKAPSTWLTAYV 1082  
Db 1006 IATYLDITTEQWETLGINRTEAVNQIVTGYAQQWYKKAHSHYAAFTNRASSSWLTAYV 1065  
QY 1083 VKVPSLAVNLIA-IDSOVLGAVKWLILEKQPDGVQFQEDAPVHCEMIGGLRNNEKDM 1141  
Db 1066 VKVFAMAAKAVAGISHBIIIOGGVRLILNRQPDGAFKNAPVLSTGMGGIQ-GABEEV 1124  
QY 1142 ALTAFLVLSIQEAKDICEOVNSLPGSITKAGDFLEANTMNLQRSYTVAIAGYALAOQMR 1201  
Db 1125 YLTAFLVILLESKTCNDVNSLDSIIKATNLLKYEKLORPYTTALTAYALAAADQ 1184  
QY 1202 LKGPLLNKFLTTAKKNRWDPKQKLYNVEATSYALLALLQLKDFDPVPPVVRWLNBOY 1261  
Db 1185 LNDRLVLMASASTGRD--HWSEYNAHNTHEG-SYALLALLKMKKFDQGPVVRWLTQDNF 1242  
QY 1262 YGGYGGSTQATFVFOALAQYQKADPDHQLNLDVSLQPSRSSKITHRIHWSASILLRS 1321  
Db 1243 YGETYGTQATVMAFALAEYEQMPTHKDLNLDITIELPDREVPVIRYRINYENALLART 1302  
QY 1322 EETKENGSTVTAAGKQGLTSVVWYHAKADQL-TCNKFDLKVITKPAETEKRPQDA 1380  
Db 1303 VETKLNQDITVTASGDGKATMTILTFYNAQLOEAKVNCNKFHLNVSVE---NIHLNANGA 1359  
QY 1381 KNTMILICTRYRGDQDATMSIIDISMTGFAFPTDDLKQLANGVDRYISKYELDKAFSD 1440  
Db 1360 KGALMLAKICTRYLGEVDSWTIIDISMLTGFLPDAEDLTELKSGVDYRYSYEVDDNMAQ 1419  
QY 1441 RNTLIIILDKVSHSDDCLAFKVOHYENVNELIOPGAVKYAYNLEESTRYHHEKEDG 1500

Db 1420 KVAVIYIYNKVSHSEBCLHPKILKHPEVGFIOGSSVKVYSYNLDEKCTKFHPDKGTG 1479  
QY 1501 KLKLCRDELRCRAEENCFTQKSDDKVTLEERLDKACEPGVDYVYKTRLVKQLSNDPDE 1560  
Db 1480 LNKICIGNVCRAGETCSSLNHQERIDVPLQIEKACETNVDYVYKTRLVKQLSNDPDE 1539  
QY 1561 YMALEQITIKSGSEVQVQORTFISPIKCBALKLEEKHYLMWGLSGDFWGEKENLSY 1620  
Db 1540 YMDVLEVIKGTDENPRAKTHQYISORCKOEAALNKVNDYDYLINGSRDLPTKDKISY 1599  
QY 1621 ITGKDTWVHPEDDEQCDSENKQCDLGAFTSMVVFQCP 1662  
Db 1600 IITKNTWIERPHEDECOEERFQKLCDDFAQPSVTLTEFQCP 1641

RESULT 11  
US-09-925-442-35  
; Sequence 35, Application US/09925442  
; Patent No. 6607897  
; GENERAL INFORMATION:  
; APPLICANT: VOGEL, CARL-WILHELM  
; KOCK, MICHAEL  
; BSEDEHORST, REINHORST  
; FRITZINGER, DAVID  
; TITLE OF INVENTION: RECOMBINANT PROCVP  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/925,442  
; FILING DATE: 10-AUG-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/017,947  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 1126-0107-0X  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1648 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 35:  
US-09-925-442-35

Query Match 48.7%; Score 4194; DB 4; Length 1648;  
Best Local Similarity 49.4%; Pred. No. 0;  
Matches 821; Conservative 324; Mismatches 481; Indels 36; Gaps 20;  
QY 9 LLLLLTHLPALGSPMYSIITPNILRESEETWVLESAHQDQGVPTVTVHDPGKGLV 68  
Db 8 LVAALLGPPGSHGALYTLITPAVLRTDTEQILVEAHGSDTPKQLIDIFVHDFPRKQKT 67  
QY 69 LSEKXVLTPTAHNGVNTFTIPANRBFKSEKKEKVTYQATFGTQV-VEKVVVLVSQ 127

68 LQTRVDMNPAAGMLVTPTIEIPA-KEVSTDSRQNVVVVQVT-GPQVLEKVKVLLSYQS 125  
 128 GYLFIOTDKTYTTCGSTVLRIRFTVNHKLLPVGRTVMNLEPBGIPVQDLSLSONQLG 187  
 126 SFLFIOTDKTYTTCGSTVLRIRFTVNHKLLPVGRTVMNLEPBGIPVQDLSLSONQLG 179  
 188 VLPFSW--DIPELVMQWIRAYENSPOQVSTEFSEFEVKEYVLPSEFEVFEVPEYF 245  
 180 -LNFEPVNLDPVLSLGTWRIWAKYHSPEN-VTAYFDRKVVLPSEFEVLPQSEKFFV 237  
 246 YNEKGLVTTIARLYGKVEGTAFLVFGIODEGQISLPESEKRIPIEDSGSEVVLGRK 305  
 238 DGNENFHSIITARYLYGEEVGAFLVFGIODEGQISLPESEKRIPIEDSGSEVVLGRK 297  
 306 VLLDGVQNPRAEDLVGKSLYVSATVILHSGSDVMQVABERGIPVTSPIYHTKTPKPKF 365  
 298 TFRS--RFPNLELNGHTLASVTVMTEGSDVMVTEGSDGHTVIVASPIYHTKTPKPKF 355  
 366 PMFPDLVMTNPDGSPAVRVPVAVOGEDTVQSLTGGDGVAKLSINTHPSOKPSITVR 425  
 366 PMFPDLVMTNPDGSPAVRVPVAVOGEDTVQSLTGGDGVAKLSINTHPSOKPSITVR 414  
 426 TKKQELSEABOATMOALPSTVGNNSNHLISVLTRELPGETLNNFLMRDRAHEA 485  
 415 TNHGDLPREARQATKSMATIAITQGGSGNYLHVATITSTEIKPGDNLVFNPNVKGANSILK 474  
 486 KIRYVYLYMKNKGLKAGROVREPQDLVLPISITTFIPSGFRVLVAYTTLIGAGORE 545  
 475 QIKFTYLLNKGLFKVGRPRDQGNLVTMLNHTPDLPSPRFAVYQV---GNNE 530  
 546 VVADSVVVDKSCVGLSVKSGSEDRQVPVQGMFTLKIEGHGAVRVVLVAVDGVFVL 605  
 531 IVADSVVVDKSCVGLSVKSGSEDRQVPVQGMFTLKIEGHGAVRVVLVAVDGVFVL 587  
 606 NKKNKLTOSKIWDVVEKIDICTPGSGKDVAGVFSAGLTFTTSSSQOQTAQRAELQCPQP 665  
 588 NKYKISQAKIWDITKSDICTAGSGQNNLGHFEDAGLALTSTNLTQKRSAAKCPQP 647  
 666 AARRR-SVQITKMDKGYX-PKELKCEGDMNPNRFSQRTFRFISLGEACKV 723  
 648 ANRRSSVLLDSSNAAAEFQODLRKCCEDVMHNPNGYTCERAKRYIOEGDACKAA 707  
 724 FLDCNRYTELRRQHASHLGLARSLNLDIIAENIVSRSEFPBSLWLNVEDL-KEPP 782  
 708 FLECCRYIKGVRENQRESEFLARDNEDGFIADSDIISRSDFPKWMLWLTOLTEEN 767  
 783 KNGISTKMLNFKLDSITTWELAVMSDKKICVADPPFVTVMQDFDIDLPLSVYRN 842  
 769 SGISKSTMSFVLRSDITWVAVSFTPTKGCICVAPPEYRVMKVFFIDLQMPYSVKN 827  
 843 EQVEIRAVLYNRQNELKVRVLLHNPFCSLATTKRRHQQTITPPKSSLSVPVYVP 902  
 828 EQVEIRAILNY-VNEDIYVRVLLHNPFCSLATTKRRHQQTITPPKSSLSVPVYVP 886  
 903 LKTLQEVKAAVYHHSISGVKSLKVPYEGIRMNKTVAVTLDPERLGRGVKEDI 962  
 887 LEQGLHDEIRASQVQEAWSGVKRLKVPYEGVQ-KSIVTVKLDPRKAGVGTOLEVI 945  
 963 PPADLSQVPTSESTRILLQGTVAQNTEDAVDAELKHLIYTPSGCGEQNMGMPTV 1022  
 946 KARKLDRVPDTEIETKIIQGDVPAQIENSIDGSLNHLIITPSGCGEQNMIRMAAPV 1005  
 1023 IAVHYLDEQWEKEGKROGALFLIKGVTOQLAFQPSAFAFAKPASTMLTAV 1082  
 1006 IATYVLDTEQWETLGINRTEAVNQIVTGAQVYKKAADHSAFTNRSASSMLTAV 1065  
 1083 VKVFLAVNLIA-IDSQVLCGAVKWLILEKQKPGVFOEDAPVHQMGGURNNEKDM 1141  
 1066 VKVFAAARWAGISHETICGVRLIILNRQOPDPAFKNAPVLSGTMOGGIQ-GAEEV 1124  
 1142 ALTAFLVLSQBXDICEQVNSLPGSTTKAGDPLEANYVNLORSYTVVALAGVAAQMR 1201  
 1125 YLTATILVALLKSTICNDYVNSLSSIKKATNVLKYEKLQRPYTTALTAYALAAADQ 1184

1202 LKGPLINKFLTTAKKNWEDFGKOLYNNVATSVALLALLQKDFDFVPPVVRWLNQRY 1261  
 1185 LNDRLVMAASGRD--HWEYNATHNIEGTSTVALLALLQKDFDFVPPVVRWLNQRY 1242  
 1262 YGGYGSTQATPMVFOALAQYKQADPHOELNLOVSIQLPSSKITHRIHWSASLLRS 1321  
 1243 YGETYQOTQATWAFQALAEYEQIOMPHKDLNLOITIELDPREVPIRYRINYENALLART 1302  
 1322 EETKEGEGFTVTAEGKGGOTLSVVTMYHAKADQGL-TCNKFDLVKTIKPAPETEKRPQDA 1380  
 1303 VETKLNQDITVTSAGDGRATWILTFYNAQIQEANCNKFLNVSVE--NIHLNAMGA 1359  
 1381 KNTMILEICTRVRGQODATMSILDISMGTGAPDTDDLKQLANGVDYRISKYELDKAFSD 1440  
 1360 KGALMLKICTRYLGEVDSWTIILDSMLTFLPDAEDLTRLSKGVDRYISRYEVNNMAQ 1419  
 1441 RNTLIYLDKSHSGSDCLAFKVOYFNVELIQGAVKVVAYYNLESCRTFRFVHEKEDG 1500  
 1420 KVAVIILANKVSHSEDECHLPKILKHFVGIQPGSVKVSYYNLDKCTKFVHPDKGTG 1479  
 1501 KLKLCRDLBCABENCFIQKSDDKVTLERLDKACEFGVDYVYKTLVKVQLSNDPDE 1560  
 1480 LNMKICIGNVCRAGETCSSLNHQBRIQIEKACETNVYVYKTLKURIEQCGNDI 1539  
 1561 YIMAIQBTIKSGSDVQVQQTFTSPICKREALKLEKHYLMWGLSSDFWGEKPNLSY 1620  
 1540 YVMDVLEVIKQGTDENPRAKTHOYISQKQCEALNKLKVNDDYLIWGSRLDLPKDKISY 1599  
 1621 IIGKDTVWEHBEDECODEENKOCODLGAFTSSMVVEGCP 1662  
 1600 IITKNYIWRWPHEDCEQEEFQKLCDDFAQFSTLTTEFGCP 1641

RESULT 12  
 US-08-447-411-76  
 ; Sequence 76, Application US/08447411  
 ; Patent No. 5773243  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FRIZINGER, DAVID C.  
 ; APPLICANT: BREDEHORST, REINHARD  
 ; APPLICANT: VOGEL, CARL-WILHELM  
 ; TITLE OF INVENTION: DNA ENCODING COBRA C3, CVF1, AND CVF2  
 ; NUMBER OF SEQUENCES: 81  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C. Jefferson Davis Highway, Suite 400  
 ; STREET: 1755 S. Arlington  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/447,411  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/043,747  
 ; FILING DATE: 07-APR-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Oblon, No. 5773243man F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 1126-101-0  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 413-3000  
 ; TELEFAX: (703) 413-2220  
 ; TELE: 248855 OPAT UR  
 ; INFORMATION FOR SEQ ID NO: 76:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1333 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-447-411-76

Query Match 41.3% Score 3559; DB 1; Length 1333;  
Best Local Similarity 51.3%; Pred. No. 0;  
Matches 683; Conservative 255; Mismatches 372; Indels 22; Gaps 13;  
QY 334 SGMVQAEAGSIPVTSPIQIHFTKTPKYPKPGMPFLMVFVTPNPDGSPAYRPV...A 390  
DB 3 SGMVQAEAGSIPVTSPIQIHFTKTPKYPKPGMPFLMVFVTPNPDGSPAYRPV...A 62  
QY 391 VQGEDTVQSLTQGGVAKLSINTHPKQKPLSITVTRKQELSEAEQATRTWQALPYSTVG 450  
DB 63 IHSECT---TLSDGTAKLFLNTPQNAQSLPTVTRNMGDLPREQAIKSNATAYQTQ 118  
QY 451 NSNNYHLSVRLTRELPGETLNVNFLMRDRAHEAKIRYTYLNMKRLKAGQVREP 510  
DB 119 GSGNYLHVAITTEIKPGDNLPTNPNVGNANSLNQIKYFTYILNKGKIFKVRGRGRGD 178  
QY 511 GQDLAVLPISITTFPSRLVAYTYLIGASQREVWADSVMDVYKDSVGVSLVVKSGQS 570  
DB 179 GENLVNWLHITPDLPSFRAVYQV---GNRIVADSVMDVYKDSVGVSLVVKSGQS 234  
QY 571 ED-RQPVGGQMTLKIEGHGARGVAVLVAVDKGVFLVKNKULQTSKIWDVVEKADIGCTP 629  
DB 235 RDNRIQMGAAKIKLEGPGAWIGLVAVDKRAEYVINDKYISQAKIWDITIEKSDPGCTA 294  
QY 630 GSGKDYAGVSDAGLFTSSGGQTAQRAELQCPQAPARRRR-SVQLTEKMDVKGY-P 687  
DB 295 GSGQNNLVGFEDAGLATTSTNLNTKQSAKCPQAPARRRRSSVLLDLSNASKAQFD 354  
QY 688 KELKCCEDGMRNPRFRSCQRTFISLGEACKVFLDCNNTYELRRQARASHGLA 747  
DB 355 QDLKCCEDGMRNPNHGTCEKREKIQEGDACKAFLPCCCHYKGIQDQDNKRESEFLA 414  
QY 748 RSNLDEDIABENIVSRSPFSSWLNVDLKEPKN-GISTKLMNIFLKDSTITWELLA 806  
DB 415 RSDPEDDLFCGNIITSRSPFSSWLNVDLKEPKN-GISTKLMNIFLKDSTITWELLA 474  
QY 807 VMSDKKGIQVAPFVWMDFFDLRLPYSVNRNEQVEIRAVLYNRQNLKRVEL 866  
DB 475 VGLSPTKIGVAPFVWMDFFDLRLPYSVNRNEQVEIRAVLYNRQNLKRVEL 533  
QY 867 LHPAFCSLATTERRHQCTITPPKSSLSVPYVIVPLKGLQEVVEKAAVYHHPISDVR 926  
DB 534 LYSAPFCSLATTERRHQCTITPPKSSLSVPYVIVPLKGLQEVVEKAAVYHHPISDVR 593  
QY 927 KSLKVPPEGIRMNKTAVATLDERLGRGVQKEDIIPADLSDQVDPTESETRILLGTP 986  
DB 594 KKLKVPPEG-EWKSIVTIELDHTKIGIGTQVELVANKLNDVPDEITETIKIIGDP 652  
QY 987 VAQMTDAVDAERKLHIVTPGCGQNNMIGMTPPTVIAVHYLDETEQWEKFLGKROCAL 1046  
DB 653 VAQMTDAVDAERKLHIVTPGCGQNNMIGMTPPTVIAVHYLDETEQWEKFLGKROCAL 712  
QY 1047 ELIKKGYTQGLAFROPSSAFAPVKEAPSTWLTAVVAVSLAVNLIA-IDSOVLCAV 1105  
DB 713 NQMTGVAQLVYKKAHSAFTNSASSSWLTAVVAVSLAVNLIA-IDSOVLCAV 772  
QY 1106 WLILEKQPGVQEDAPVHQBMSIGLNNKEMALTAFLVLSLOAKDICSEQVNSL 1165  
DB 773 WLILNRQTDGVFRENAPVLFMTMOGQIQ-GAEPGSLTAFILVALLSRSICNAYNIL 831  
QY 1166 PGSITKAGDPLEANYNLQSVYVAVAGYALQMGRLKGPLLNKFLTTAKDKNWDPGK 1225  
DB 832 DSSISKATDYLKYEKLRPYTALTAYALAAAEELNDRV-LMAASTGRNWEFPA 889  
QY 1226 QLYNVEATSVALLALLQLQDFDVPVYVWLNQRYGGYGGTQATFWFOALAQYOKD 1285

DB 890 HTHNIEGTSVALLALLKMKKFEAGPVQVWLIDQQYVGGTYGQTQATVMPFOALAEYBQ 949  
QY 1286 APDQELNLDVSLQPSRSSKITHRIHWSASALLSEETKENEFTVTAEGQOGLSVV 1345  
DB 950 MPTHKDLNLDITIELPDREPIRYRINYENALIACTVETKLNEDFTVSASGDKATMIL 1009  
QY 1346 TWYHAKAK-DQLTCNKFDLKVTIKPAPETEKRPDOKAKNTMILEICTRYRGDODATMSILD 1404  
DB 1010 TVYNAQLREDAVNCNPHLDVSVENQLNLKKAAGAKGALKIKICTRYLGEVDSVNTIIL 1069  
QY 1405 ISMTGTAPDITDQLKLANGVDRYISKYELDKAFSDRNLTLLIYLDKVSHEBDDCLAFKVH 1464  
DB 1070 VSMLTGVPDTEDLTKSGVDRIYSIMFEINNMAQKGTVIIYLDKVSHEBDDCLAFKVH 1129  
QY 1465 QYFNVLTIQGAVKAVYAYNLEESCTRFYHPEKEDGKLNKCRDLBCAEEFCIQKSD 1524  
DB 1130 KHFEVGHQFGSVKYYSYNLDKCTKIYHPDEATGLANKICVGNVCRCAEETCSLLNQ 1189  
QY 1525 DKVTLEBRLDKACEPGVDVYVYKTRLVKVLQSLNDFDEYIMAEIQTIKSGSDEVQVQOQRTF 1584  
DB 1190 KNTVTRQLRIQKAFDPNDVYVYKTRLVKVLQSLNDFDEYIMAEIQTIKSGSDEVQVQOQRTF 1249  
QY 1585 ISPTKCREALKLEKHYLMWGLSSDFWGEKKNLSYIIIGKDTWVSHWPEDEECQDEENQK 1644  
DB 1250 VSQRKCEALNLMVNDYLLNGPSSDLMPMKOKISYLIITNTWIERWPHEDKQCEEFQK 1309  
QY 1645 QCQDLGAPTESM 1656  
DB 1310 LCDDFALFSYAM 1321  
RESULT 13  
US-08-662-227-34  
Sequence 34; Application US/08662227  
Patent No. 5922320  
GENERAL INFORMATION:  
APPLICANT: VOGEL, CARL-WILHELM  
APPLICANT: BREDEHORST, REINHORST  
APPLICANT: KOCK, MICHAEL  
APPLICANT: FRITZINGER, DAVID  
TITLE OF INVENTION: RECOMBINANT PROCVF  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/662,227  
FILING DATE: 14-JUN-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 1126-0107-0X  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1333 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein



Query Match 41.1%; Score 3536; DB 3; Length 1333;  
Best Local Similarity 51.2%; Pred. No. 0;  
Matches 683; Conservative 254; Mismatches 372; Indels 24; Gaps 15;

QY 334 SGMVQAESESGIPVTSPIQIHTKTPKYPKPGMPFDMVFNTPDGPSPAYRPV...A 390  
DB 3 SGMVQAESESGIHTVTSPIQIHTKTPKYPKPGMPFDMVFNTPDGPSPAYRPV...SEA 62  
QY 391 VQSDTVOISITQDGVAKLSINHPKOKPLSTIVTKKQELSEABOARTMOALPYSTVG 450  
DB 63 IHSEGT---TLDGAKLFLNTPQNAQSLPITVTRNGDLPREBQAISKMTATAYOTQ 118  
QY 451 NSNNYHLVRLTELPAGETLNVNLLRMDRAHEAKIRYIYVILNKKGLLAKGQVREP 510  
DB 119 GSGNYLVAITSTEIKPGNLPFNFNVRGNANSLQIKYFTYIILNKKIKFKVGRHRGD 178  
QY 511 GQDLVLPSTITTFPSFRLVAYYTLIGASQREV-VADSVVVDKSCVGSVWKSQ 569  
DB 179 G-NLVTNMLHITDLPISFRFVAYQV---GNNEIEVADSVVVDKTCGTLLVKGAT 233  
QY 570 SED-RQVPQOQMTLKIEGDHGRVVLVAVRGVFLNKKGLTOSKIWDVVEKADTGT 628  
DB 234 SRDNRIOQPAAMKIKLEGDPGAWIGLVADRAEYVLDNKKYISQAKIWDTIERSDFGT 293  
QY 629 PGSGKYAGVFS DAGLTFITSSSQQAQAEIQCQPAARRRR-SVQLTEKRMKVKGY- 686  
DB 294 AGSQNNLGVFEDAGLALITSTNLTKQSAKACQPAARRRRSSVLLDSNASKAAQFQ 353  
QY 687 PKELKCCBGMENPMRSCQARTFISLGACKVFLDCNVTITELRROHARASHLGL 746  
DB 354 QDLURKCCBGMENPMHGTCEKERYIQEGDACAFLCCHYIKGIDDKKESELEL 413  
QY 747 ARSLDEDIIAENINYSRSEFPELWVVEDLKEPPKQ-GISTKLMIFFLKDSTTTWEIL 805  
DB 414 ARSDFEDDLFEGENITSRSDFPESLWLMELQSEHSPKSGISKIYVPLRDSITTWELL 473  
QY 806 AVNSDKKGTGVADPEVTVQDFFDLPLPYSVVNEQVEITRAVLNRYONQELKVRVE 865  
DB 474 AVGLSPFKGICVAPETVITWMDFFDLPLPYSVVNEQVEITRAVLNRYONQELKVRVE 532  
QY 866 LLNPAFCSLIATKRHOQTITIPPKSSLSVPYIVPLKTLQEVKAAVYHFFISDGV 925  
DB 533 LLSPAFCSASTESQRYEOLPIKALSSRAVSFVIVPLEQGLHDVEVTASVOGLMGDGV 592  
QY 926 RKSILVYVEGIRNKNKTVARTLDPERLREGVQKEDIIPADLSQVDPDTSETRILQGT 985  
DB 593 KKKLVYVEG-ENKSVITIELDPHTKIGGTQVQLVANKLNDKVPDTEITKITIQGD 651  
QY 986 PVAQMTEDAVDAERLKHILVTPSGCEQNMIGMTPTVIAVHYLDBTEQWEXFLEKROGA 1045  
DB 652 PVAQTIENSIDGSKLNLHILITPFCCEQNMIRMTAPVIATYVLDITQWETLGNRTEA 711  
QY 1046 LELIKGYTQOLAFROPSAFAFVRAKSTWLTAVVVKVSLANLIA-IDSOVLGAV 1104  
DB 712 VQQLMTGYAQLVYKADHSYAFNTSASSNLTAVVVKIIPALAKIVKINDINHIIVCGM 771  
QY 1105 KWILEKQKDPGQFQDADPVIHOEMIGGJRNNEKDMALTAFLVLSIOEAKDICEOVNS 1164  
DB 772 RWLLNQRDGDVFRENAPVLFQTMGGIQ-GAEPEGLTAPILVALLESRSICNAVINI 830  
QY 1165 LPSGSIKAGDFLEANTYMNQORSYTVAIAGYALQAGMLKPLNKLFTTAKDKNRWEDPG 1224  
DB 831 LDSSISKATDYLKKYKELQRPYTTALTAYALAAERLNDRV--LMAASTGRNRWEPN 888  
QY 1225 KQLYNVEATSYALLALQLKDFPVPVVRNLEBRYGGVGSTOATFWVFOALAYQK 1284  
DB 889 ARTHNIEGTSYALLALKKKVEAGVAVVQVLDIQYGGTYGQATVMFMFOALAEYBI 948  
QY 1285 DAPDQELNDVSLQFSSSKITHRIHWESASLSRSEETKENEGFTVTAEGKQOGLTS 1344  
DB 949 QMPTHKDLNDITIELPDRVPIRYINYNALLAQVETKLNEDFTVSASGDCKATWTI 1008  
QY 1345 VTMVHAKAK-DQLTCNKFDLKVTKIPATEKRPQDANKTMLIEICTRYRQDQATWSIL 1403

DB 1009 LTVNAQLREDANVCNKFHLDVENVQNLNKEAKGALKKICTRYLSEVSTWTLII 1068  
QY 1404 DISMXTGAPDITDOLKANGVDRIYSKYELDKAFSDRNTLIYLDKVSHEDEDDCLAFKV 1463  
DB 1069 DVSMLTGFEVDTEDTLRLSGVDRIY:SMFEINNNMAQKGTIYLDKVSHEDEDECLAFKI 1128  
QY 1464 HQYNVBLIOPGAVKYVAYYNLEESCTRFVHPHEKEDGKLNKCRDELRCRCAEENCFIOKS 1523  
DB 1129 LKHEVGFIOGSKVYVYNNLEKCTKIYHPDEATGLLNKICVGNVCRCAEETCSLLNQ 1168  
QY 1524 DDKTLBERLDKACEPOVDYVYKTRLVKVLQSLNDFEYIMAEQTIKSGSDEVQVQOQT 1583  
DB 1189 QKNTROLRIQKAPDPNDVYVYKTLRLTEKXGNDIYVMDVLEVLKQGTQNOQOVKVRQ 1248  
QY 1584 FISPIKREALKLEEKHYLMWGLSDFWGEKPNLSYIYIKDTWVHWPEDDECOBENQ 1643  
DB 1249 YVSQKQCEALNLMVNDYLLWGPSSDLWPMKDKISYLITKNTWIERPHEDKQCEEBFQ 1308  
QY 1644 KQCODLCAFTSM 1656  
DB 1309 KLCDDFALFSYAM 1321

RESULT 15  
US-09-925-442-34  
Sequence 34, Application US/09925442  
Patent No. 6607897  
GENERAL INFORMATION:  
APPLICANT: VOGEL, CARL-WILHELM  
BREDEHORST, REINHORST  
KOCK, MICHAEL  
FRITZINGER, DAVID  
TITLE OF INVENTION: RECOMBINANT PROCVF  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/925,442  
FILING DATE: 10-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/017,947  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 1126-0107-0X  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1333 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 34:  
US-09-925-442-34

Query Match 41.1%; Score 3536; DB 4; Length 1333;



**This Page Blank (uspto)**